Predictive analysis of dynamical systems: combining discrete and continuous formalisms

Madalena Chaves

UNIVERSITÉ DE NICE SOPHIA ANTIPOLIS

Mémoire
présenté pour obtenir le diplôme d'
Habilitation à diriger des recherches
en
Sciences & Technologies de l'Information et de la Communication
par

Madalena CHAVES

Predictive analysis of dynamical systems: combining discrete and continuous formalisms

Date de soutenance: October 24, 2013

Rapporteurs:	David ANGELI	Imperial College/University of Florence
	Alexander BOCKMAYR	Freie Universität Berlin (jury)
	Leon GLASS	McGill University (président du jury)
Examinateurs:	Jean-Luc GOUZÉ	INRIA Sophia Antipolis (jury)
	Christophe PRIEUR	CNRS Gipsa-lab Grenoble (jury)
	Thomas SAUTER	University of Luxemburg (jury)

Abstract

The mathematical analysis of dynamical systems covers a wide range of challenging problems related to the time evolution, transient and asymptotic behavior, or regulation and control of physical systems. A large part of my work has been motivated by new mathematical questions arising from biological systems, especially signaling and genetic regulatory networks, where the classical methods usually don't directly apply. Problems include parameter estimation, robustness of the system, model reduction, or model assembly from smaller modules, or control of a system towards a desired state. Although many different formalisms and methodologies can be used to study these problems, in the past decade my work has focused on discrete and hybrid modeling frameworks with the goal of developing intuitive, computationally amenable, and mathematically rigorous, methods of analysis.

Discrete (and, in particular, Boolean) models involve a high degree of abstraction and provide a qualitative description of the systems' dynamics. Such models are often suitable to represent the known interactions in gene regulatory networks and their advantage is that a large range of theoretical analysis tools are available using, for instance, graph theoretical concepts. Hybrid (piecewise affine) models have discontinuous vector fields but provide a continuous and more quantitative description of the dynamics. These systems can be analytically studied in each region of an appropriate partition of the state space, and the full solution given as a concatenation of the solutions in each region.

Here, I will introduce the two formalisms and then, using several examples, illustrate how a combination of different formalisms permits comparison of results, as well as gaining quantitative knowledge and predictive power on a biological system, through the use of complementary mathematical methods.

Contents

Chapter 1

Introduction

The mathematical analysis of dynamical systems is a vast area of research, encompassing many challenging, as well as fascinating, problems related to the temporal evolution, behavior, or regulation and control of physical systems. Many different formalisms and methodologies can be used to study these problems, such as continuous ordinary or partial differential equations' models [14, 44, 37], discrete or hybrid frameworks [61, 62, 29, 30], and also stochastic elements [41] (see also [40], for a review).

In the past decade, a large part of my research work has been on mathematical analysis of models motivated by biological systems, especially signaling and genetic regulatory networks. Due to their complexity and size, (continuous) models of biological regulatory systems tend to be analyzed by numerical simulations [44], which provide an invaluable means for studying the systems, but also involve numerical approximations (hence liable to numerical errors), which often do not capture the whole range of dynamical behaviors of the given system, as only a finite number of parameter sets can be tested. In contrast, discrete models provide a (sometimes very) abstract representation of physical mechanisms, but many exact and rigorous computational tools are available [12, 32, 19] that can be used to study such system without incurring in numerical approximations.

My goal has been the development of mathematically rigorous methods of analysis with the objective of contributing to study, understand, and eventually control the behavior of biological regulatory systems. This Introduction will motivate my choice of mathematical modeling frameworks and give a brief description of the main two formalisms used in my work by quickly recalling their most relevant properties. I will also underscore the links between the two formalisms, and illustrate their suitableness and complementarity for addressing questions inspired by biological systems. [1]

1.1 Mathematical models for biological networks

Biological systems have particular constraints, and often give rise to new mathematical problems for which classical methods do not apply [60]. In general, the variables in a mathematical model will denote the concentrations of a set of molecular species (which can be proteins, messenger RNAs, biochemical compounds) involved in a network of interactions. There are typically many species involved, giving rise to large networks (easily reaching tens to hundreds of components), where the same species may be involved in more than one system, which may lead to complex networks of interactions and cross-talks [44]. Despite their complexity, as well as cell-to-cell variability, biological regulatory networks are known to exhibit rather robust properties in response to external or internal stimuli [43]. The analysis of such networks and the connection between

[1] NOTE ON CITATIONS: the citations to papers where I am a co-author are preceded by the letter C, and can be found in Appendix A. My list of publications is numbered in reverse chronological order and separated into three general categories: journals and book chapters, conference proceedings, and technical reports (for instance, [C7] is a journal article and can be found in Appendix A.1). All other citations are numbered in alphabetical order and listed in the Bibliography section.

the structure or topology of interactions and the robustness properties of the dynamical outcomes are common problems in molecular biology [43, 66, 2].

On another level, there are the measurements and external actions on the systems (outputs and inputs, in control theoretical language), which greatly depend on, and are limited by, the experimental techniques available. The type of data and information can vary from system to system, affecting model construction and mathematical representation.

A number of questions have arisen in the context of "synthetic biology", a recent research discipline which aims at designing and constructing cellular networks from basic biochemical components. Problems related to the coupling of individual components, their regulation and control to produce a desired dynamical outcome can be studied through mathematical models [7, 27, 18, 3, 64, 67].

In my view, a mathematical model is useful if it can be compared to data, reproduces some properties of the system, and leads to predictions, such as qualitative behaviors, orders of magnitude of some parameters, or hints to control strategies. To construct and analyze a mathematical model of a biological system – and in the process gain insight into the system's properties – new tools and techniques need to be developed, apt to deal with the available data and the questions concerning the system. This search for appropriate methodology has lead me to define and develop a (hopefully) coherent research direction.

Data constraints and mathematical models. To model and analyze a genetic network, one of the first aspects to be considered is the type of data available for the system. Biochemical experimental techniques have rapidly evolved in the past decade, and there are currently many different types of measurements available, from qualitative micro-array data (presence or absence of a DNA motif) to more quantitative data with reporter genes or fluorescent fusion proteins (see [44] for a review). The latter can have relatively high frequency sampling (relative to the time scale of the phenomenon) and yield quite smooth data. A second aspect is the number of components that are known to be part of the system or that one wishes to consider as part of the process. Among these, there will be those components that can be measured, and those that one wishes to control or follow in some way (see also [67]).

I have come to realize that the choice of model formulation is a central aspect of the analysis, as it will determine the type of methods that can be used, as well as the predictive power of the model. One of my objectives has been to construct models that help to extract useful information from the available data, but also carrying the theoretical analysis as far as possible.

Two modeling frameworks have since imposed themselves throughout my work, due to their suitability to handle the data available on biological regulatory systems, their amenability to theoretical analysis and implementability as computational tools. Here, I will thus focus on Boolean and piecewise affine formalisms, which I believe have proved and will continue to prove useful for the analysis of dynamical systems. Boolean models describe the topology of the network of interactions through logical functions (with no parameters) and provide intuition on the dynamical behavior of the system [30, 62]. Piecewise affine (PWA) systems provide a simplified continuous framework, as they describe synthesis and degradation terms with only a reduced number of parameters [29, 28, 10].

Model construction. The actual model construction presents several challenges, such as how to mathematically interpret the effect of one variable on another, or more generally the effect of two or more variables on another (see below), and then choosing a function that describes it, as the biological data does not always help here. Detailed continuous modeling approaches such as mass-action kinetics [37] generally provide a precise stoichiometric description of all the *elementary reactions* that happen in a biochemical system. However, they also require a large number of parameters, which in turn require a large number of experiments, many of which are often difficult to perform, or too expensive, or even not possible. After my PhD thesis (where I studied a class of mass-action kinetics' models), I have explored other modeling approaches that would not need so

many (unknown) parameters, and could be more easily compared to the available measurements. If continuous, mass-action, models are at an extreme end of the modeling scale, then Boolean models are at the other extreme as they involve no parameters and are formulated according to logical rules using only a pair of states, true/false or 0/1, for each variable. At an intermediate level of the modeling scale, piecewise affine systems are a continuous but abstract framework, based on a finite partition of the state space and involving only a restricted number of parameters. It has been observed that activity of a gene follows a sigmoid-shaped curve [70, 48], reasonably represented by Hill functions with exponent larger than 2. In fact, PWA systems can be said to use approximations of Hill functions as the exponents tend to infinity.

It is always difficult to simplify a chain of two or more events and represent them by a single mathematical term. There are no correct nor unique answers, but an advantage (or a disadvantage, depending on the point of view...) of the Boolean and piecewise affine formalisms, is that these simplified representations can be reduced to a small number of choices and mathematical analysis can then help to discriminate between various possible model versions of the same system. Another advantage is the fact that various computational and algorithmic tools are available for the study of PWA systems and Boolean networks. These tools provide an analysis of the dynamics of the system, its asymptotic behavior (steady states, periodic solutions) and their stability. In many cases, this information can then be used to guide the construction and analysis of a continuous, more detailed, model of the system.

A combination of different formalisms for a predictive analysis. Each of these frameworks has advantages and limitations, but both have allowed for the development of mathematically rigorous methods. While they can be used independently, these two formalisms can be related and coupled to complement and enrich each other, thus increasing the predictive power of a model.

There are several ways to combine the two formalisms. For instance, the logical description contained in a Boolean model can be used to construct a piecewise affine model. Similarly, the state spaces (or the state transition diagrams) of the two models can be related, and the qualitative information contained in the Boolean model can provide intuition on the continuous dynamics. Conversely, the parameters of PWA systems can be combined to (re-)introduce quantitative aspects into the Boolean model, such as the notion of relative timescales of different physical phenomena, or the probability of a given biochemical reaction or event.

Therefore, the methods described here yield both qualitative and quantitative aspects to characterize the robustness, variability, and dynamical behavior of biological regulatory systems.

I will next give a brief description of the main properties of each of these modeling formalisms, but first introduce an example of a (simplified) genetic regulatory network, which will be used below to illustrate model construction.

An example. Consider the transcription factor NFκB which contributes to the production of its own inhibitor, IκB, a process that was first observed and reported by Hoffmann et al. [36]. The following phenomena have been observed:

(a) cytoplasmic NFκB (NFκB$_{\mathrm{cyt}}$) is sequestered by the molecules of IκB (forming a bound complex);

(b) upon external signaling, IκB is phosphorylated and degraded through the action of an IκB kinase;

(c) this frees the molecules of NFκB which can now translocate to the nucleus (NFκB$_{\mathrm{nuc}}$) where they activate transcription of the gene encoding for IκB;

(d) when bound to IκB, NFκB cannot activate transcription of the gene.

These steps are schematically represented in Fig. 1.1, where a normal (resp., blunt) arrow represents a positive (resp., negative) effect from the originating to the target species. Several other species are known to be involved

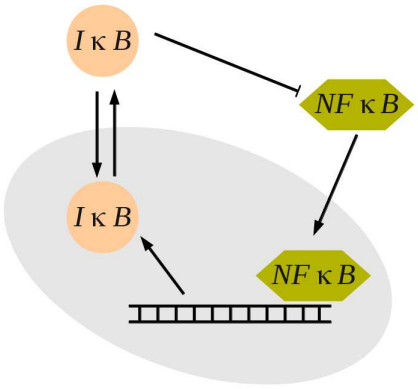

Figure 1.1: Schematic representation of the main interactions between transcription factor NFκB and its own inhibitor, IκB. The shaded area represents cell nucleus and the stripped band represents DNA.

in the interactions between NFκB and IκB (notably, other genes whose transcription is also triggered by NFκB). A model for this extended network was developed in [47], which includes the interactions pictured in Fig. 1.1.

Many simplifications have been assumed to draw this diagram, for instance, all the transcription and translation steps for IκB protein synthesis are omitted (binding of transcription factor to the gene, mRNA formation, etc.): these steps are represented by the positive arrow NFκB$_{\text{nuc}}\rightarrow$IκB, which means "whenever there is an abundance of NFκB, it will lead to an increase in the concentration of protein IκB". Likewise, the formation of the [NFκB-IκB] complex is not detailed, but simply represented in the diagram by a blunt arrow that means "whenever there is an abundance of IκB, it will lead to a decrease in the concentration of NFκB." This schematic regulatory network will be used in Sections 1.2-1.4 to illustrate the various formalisms.

1.2 Classical models: ordinary differential equations

A large range of ordinary differential equations (ODEs) models are regularly used for describing biological systems [41, 35, 14, 5], each of them allowing different levels of theoretical analysis. One of the most classical frameworks for modeling biochemical networks consists of systems in which the vector fields are polynomial. They are obtained by assuming that each biochemical reaction or event follows the *law of mass-action* [37, 35] in an "ideal mixture" (i.e., assuming that all the species are homogeneously distributed, at a fixed temperature, and in a given time invariant volume). The law of mass-action says that two (or more) species A and B may combine to produce a new species (called the "complex", C) and the rate of formation of C is proportional to the amounts of A and B present in the medium: k_1AB. The dynamical properties of some classes of mass-action systems can be studied in great detail as in the case of zero-deficiency networks[24, 23] or [C24], [C22] – however, a review of such topics is not the aim of this section.

In general, since the variables in a mathematical model denote the concentrations of molecular species, an immediate constraint is that the mathematical system and its solutions should exist and remain non-negative for all times. Given a network with n species, the variables will be $x \in \mathbb{R}^n_{\geq 0}$ and the model

$$\dot{x} = f(x), \quad x(0) = x_0$$

where f is typically Lipschitz continuous (or piecewise Lipschitz, in Section 1.3), so that existence and uniqueness of solutions is guaranteed [59]. In addition, in order to have invariance of the non-negative orthant $\mathbb{R}^n_{\geq 0}$,

the vector fields should also satisfy

$$x_i = 0 \;\Rightarrow\; f_i(x) \geq 0, \quad \forall i = 1, \ldots, n.$$

These two aspects are usually taken into consideration during the construction of a model.

Example A mass-action kinetics type formalism is used in [47] to model a more detailed version of the network shown in Fig. 1.1. To simplify the notation I will write:

$$x_a = \text{NF}\kappa\text{B}_{\text{cyt}}, \;\; x_b = \text{I}\kappa\text{B}_{\text{cyt}}, \;\; x_r = \text{NF}\kappa\text{B}_{\text{nuc}}, \;\; x_s = \text{I}\kappa\text{B}_{\text{nuc}}$$

and also the bound complexes:

$$x_c = [\text{I}\kappa\text{B}_{\text{cyt}}\text{-NF}\kappa\text{B}_{\text{cyt}}], \quad x_d = [\text{I}\kappa\text{B}_{\text{nuc}}\text{-NF}\kappa\text{B}_{\text{nuc}}], \quad x_k = [x_c\text{-IKK}]$$

where IKK is another species (not shown in Fig. 1.1), which contributes to liberate $\text{NF}\kappa\text{B}_{\text{cyt}}$ from its inhibitor. Using mass-action kinetics, the following system of differential equations is obtained:

$$\begin{aligned}
\dot{x}_a &= k_2 x_c - k_1 x_a x_b - d_1 x_a + k_6 x_k \\
\dot{x}_b &= k_0 - k_1 x_a x_b - k_5 x_b x_c - d_2 x_b + d_4 x_s \\
\dot{x}_r &= \alpha d_1 x_a - k_3 x_r x_s \\
\dot{x}_s &= \alpha d_2 x_b - k_3 x_r x_s - \alpha d_4 x_s \\
\dot{x}_c &= k_1 x_a x_b - k_2 x_c - k_5 x_b x_c + d_3 x_d \\
\dot{x}_d &= k_3 x_r x_s - \alpha d_3 x_d \\
\dot{x}_k &= k_5 x_b x_c - k_6 x_k
\end{aligned} \tag{1.1}$$

The equation for cytoplasmic NFκB ($\dot{x}_a$) contains two "production" terms due to dissociation of the complexes ($k_2 x_c$ and $k_3 x_d$), and two loss terms due to binding to its inhibitor ($k_1 x_a x_b$) and transfer to the nucleus ($d_1 x_a$). The equation for nuclear NFκB ($\dot{x}_r$) has one production term corresponding to the transfer from the cytoplasm ($\alpha d_1 x_a$) and one loss term due to binding to its inhibitor ($k_3 x_r x_s$). In the equations for IκB there is one production term due to translation of IκB mRNA (denoted k_0), there is a two-way transfer of IκB between the cytoplasm and the nucleus (terms $d_2 x_b$, $d_4 x_s$), and note that IκB is not recovered on the dissociation of complex x_c. The equations for the complexes are similarly constructed.

Model reduction This model accounts for many details such as complex formation or various forms of the same species. Some mathematical techniques are available to study mass-action models [35, 24, 23] but, on the other hand, model (1.1) involves many parameters, most of which are unknown or difficult to estimate. So, one often looks for ways of simplifying the model and reducing it to a more compact, but mathematically more tractable, model. A general simplifying assumption is to consider that some phenomena happen on a faster timescale relative to others (for instance, complex formation is faster than transcription or translation [2]). This leads to the hypothesis that some equations will be at *quasi steady state* (see, for instance, [14]), which can be justified by identifying a small parameter (or combination of parameters), ε and re-writing the equations in the form:

$$\begin{aligned}
\dot{x}_1 &= \frac{1}{\varepsilon} f_1(x_1, x_2) \\
\dot{x}_2 &= f_2(x_1, x_2).
\end{aligned}$$

Roughly speaking, $\varepsilon \dot{x}_1 = f_1(x_1, x_2) \approx 0$ (if ε is "small" and f_1 not "too large"), yields an algebraic equation that defines x_1 as a function of x_2, $x_1 = g(x_2)$. Thus, x_1 is called the "fast" variable, as it vector field or

velocity has larger values relative to those of variable x_2. Elimination of the x_1 equation as well as substitution of x_1 as a function of x_2 leads to a reduced model for the dynamics of x_2: $\dot{x}_2 = f_2(g(x_2), x_2)$. There are however, several mathematical conditions that should be verified in order to establish the range of validity of this model reduction: these are known as Tikhonov's Theorem (see, for instance, [42], and a brief summary in the introductory book chapter [C5]; another application of this technique can be found in [C10]).

Example (continued) Assuming that complex formation is fast, one can use the quasi-steady state assumption to approximate $\dot{x}_i \approx 0$ ($i \in \{c, d, k\}$) in model (1.1). Other simplifying assumptions can be introduced, depending on the system. In this example, I will also consider that:

- there is only one pool of IκB and merge the corresponding cytoplasmic and nuclear forms, $x_b + x_s \rightsquigarrow x_b$;
- for each variable, all degradation or loss terms are grouped into a single linear term.
- for each variable, all production terms are grouped into a phenomenological (positive or negative) activation term that represents the "overall effect".

Following these guidelines, an alternative, and more schematic, three-dimensional model can be written with linear degradation terms and production or *activity functions* (denoted h^j):

$$\dot{x}_a = k_a h^a(x_b) - \gamma_a x_a \tag{1.2}$$
$$\dot{x}_b = k_b h^b(x_r) - \gamma_b x_b \tag{1.3}$$
$$\dot{x}_r = k_{ra} h^{ra}(x_a) + k_{rb} h^{rb}(x_b) - \gamma_r x_r \tag{1.4}$$

Note that the activity functions are in agreement with the arrows pictured in Fig. 1.1; they could have several mathematical forms, but h^a and h^{rb} should be decreasing functions of their argument (since x_b inhibits both x_a and x_r), while h^{ra} and h^b should be increasing functions (see also [C19] for a general idea). Since each activity function generally represents a sequence of elementary events, typical forms for h^j in signaling and genetic networks can be obtained by application of hypotheses on the timescales of the events and appropriate assumptions on the parameters (see, for instance, [14] or [2]; see also the introductory book chapter [C5]):

$$h^+(x, \theta, m) = \frac{x^m}{\theta^m + x^m}, \quad h^-(x, \theta, m) = \frac{\theta^m}{\theta^m + x^m},$$

where $m \geq 1$ and $\theta > 0$ are positive real numbers. It is clear that h^+ (resp., h^-) is increasing (resp., decreasing). The number θ represents a threshold above (resp., below) which the effect of variable x is very strong. For instance, in the x_b equation, there will be a strong transcription/translation of x_b mRNA once x_r is above a certain value θ_{a3}. These sigmoid functions (also called Hill functions) are known to fit well to activity data [48], leading to exponents of the order 2-4 [70], and are often used to model genetic regulatory networks [18, 27].

Theoretical analysis of systems of ODEs is very difficult, especially as the number of variables increases. Even in this example with three variables, it is difficult (or impossible) to obtain explicit expressions for the steady states. Besides, mass-action kinetics, there are other classes of systems for which some tools exist (for instance, monotone systems [4]) but, in general, their stability can be studied only locally and some qualitative properties can be established (see [14]). For an idea of the dynamical behavior, it is therefore usual to rely on numerical simulations (as in [47]). Another possible approach is to consider an approximation of system (1.2), based on the form of its vector field: if m is very large, the functions $h^\pm$ become close to Heaviside (or "step") functions, and system (1.2) can be approximated by a piecewise linear system. This fits into a useful framework that will be discussed next.

1.3 Piecewise affine models

Piecewise affine (PWA) models also consist on systems of differential equations, but the vector fields have (finitely many) points of discontinuity [34, 10]. Briefly, the vector field may take different expressions in different regions of the state space. However, these expressions must be affine or linear in each variable (no quadratic terms are allowed).

Intuitively, a piecewise affine model can also be obtained as a limiting case of a classical system of ODEs whose vector fields is given in terms of Hill functions, by letting the exponents tend to infinity (see, for instance, [C17]). In this case, each Hill function becomes a step function with the discontinuity at the point $x = \theta$.

$$\lim_{m\to\infty} h^+(x,\theta,m) = s^+(x,\theta) = \begin{cases} 0, & \text{if } x < \theta \\ 1, & \text{if } x > \theta. \end{cases}$$

Note that the function $s^+(x,\theta)$ remains undefined at $x = \theta$, which are the points (or hyper-surfaces) of discontinuity of the vector field. Similarly, the decreasing Hill function yields a decreasing step function: $s^-(x,\theta) = 1 - s^+(x,\theta)$.

In the context of genetic networks, L. Glass introduced a class of PWA models where the vector field is obtained from logical activity functions [30, 29, 16, 15]. Alternatively, a PWA model can also be constructed directly from the available biological knowledge (such as in [52]). More applications of PWA systems to biological systems can be found in [C21], [6]. The main properties of PWA systems and briefly summarized next.

Regular domains, focal points, and solutions As above, I will consider variables $x = (x_1, \ldots, x_n)' \in \mathbb{R}^n_{\geq 0}$, which can be assumed to evolve in a bounded subset of $\mathbb{R}^n_{\geq 0}$, $R = [0, \text{Max}_1] \times \cdots [0, \text{Max}_n]$, as will become clear below. In general, each variable x_i may have a finite number of activity/inactivity levels (or thresholds), describing its interaction with the other variables in the network:

$$0 = \theta_{i0} < \theta_{i1} < \cdots < \theta_{i,p_i} = \text{Max}_i < \infty.$$

These thresholds define rectangular regions which form a partition of the positive orthant:

$$\mathcal{D} = \{(\theta_{1,r_1}, \theta_{1,r_1+1}) \times \cdots \times (\theta_{n,r_n}, \theta_{n,r_n+1}) : r_i \in \{0, p_i - 1\}\}.$$

The rectangular regions in $\mathcal{D}$ are called *regular domains*, while their boundaries are hyper-planes characterized by the fact that some of the variables are at a threshold and are called *switching domains*, D^s. A convenient way to label the regular domains is to use the indexes of the first threshold of each variable:

$$(\theta_{1,r_1}, \theta_{1,r_1+1}) \times \cdots \times (\theta_{n,r_n}, \theta_{n,r_n+1}) := \text{“}r_1, \ldots, r_n\text{,”} \tag{1.5}$$

so that, if $n = 3$, the domain “120” corresponds to the 3-dimensional region (see also example below):

$$\text{“120”} \;\Leftrightarrow\; x_1 \in (\theta_{11}, \theta_{12}),\; x_2 \in (\theta_{22}, \theta_{23}),\; x_3 \in (\theta_{30}, \theta_{31}).$$

The general formulation of a PWA system is as follows:

$$\dot{x} = f(x) - \Gamma x$$

where $\Gamma = \text{diag}(\gamma_1, \ldots, \gamma_n)$ and the vector field $f : \mathbb{R}^n_{\geq 0} \to \mathbb{R}^n_{\geq 0}$ takes different forms depending on the region of the state space:

$$f(x) = f^D(x), \quad \forall x \in D, \;\; D \in \mathcal{D}.$$

Throughout my work, the PWA are characterized by vector fields which are constant in each region (i.e., f^D is a constant), which implies that the equations are decoupled, and the solution can be explicit computed for all $x \in D$:

$$\dot{x}_i \quad = \quad f_i^D - \gamma_i x_i, \quad i = 1, \ldots, n$$

with solution

$$x_i^D(t) = (x_{i0} - M_i^D)e^{-\gamma_i t} + M_i^D, \quad \text{where } M_i^D = \frac{f_i^D}{\gamma_i}.$$

For each domain D, it is clear that solutions will evolve towards $\phi^D = (M_1^D, \cdots, M_n^D)$, called the *focal point of* D. Since there are only a finite number of thresholds and the f_i^D are constant, (re-)define $\text{Max}_i = \max_D\{M_i^D\}$. Then the set R is an invariant region for system and one may consider that $x(t) \in R$ for all $t \geq 0$. Each focal point ϕ^D may be contained inside or outside the domain D. In the latter case, solutions eventually leave the domain to enter another one, and the system switches to another vector field. In the former case, the domain is invariant, and the focal point becomes a true fixed point (see also [10]). If the vector fields in two adjacent domains do not have opposite orientations, the solution can be "normally" continued by concatenating the solutions in the two domains. Otherwise, the vector field has to be defined as a differential inclusion along the switching surface shared by the two domains, and a solution can still be constructed in the sense of Filippov ([26]; see also [33]).

Switching domains and sliding modes Note that the vector fields of the PWA systems are undefined at the points (or hyper-surfaces) of discontinuity. Nevertheless, solutions can still be continuously defined, using a construction due to Filippov [26], as follows. On a switching domain D^s, the system is defined as a differential inclusion:

$$\dot{x} \in H(x) = \overline{\text{co}}\{f^D(x) - \Gamma x : \ D^s \in \partial D\}, \text{ if } x \in D^s$$

where $\overline{\text{co}}$ denotes the closed convex hull of the vectors and ∂D contains all switching domains adjacent to D. For example, if two domains D_a and D_b share the face D^s characterized by $x_1 = \theta_1$, then *for all* $x \in D^s$, the solutions may satisfy:

$$\dot{x} = \alpha f^{D_a}(x) + (1 - \alpha) f^{D_b}(x) - \Gamma x.$$

for any $0 \leq \alpha \leq 1$. If $f_1^{D_a} - \gamma_1\theta_1$ and $f_1^{D_b} - \gamma_1\theta_1$ have the same sign, then there is a "natural continuation" of the trajectory as it evolves from D_a towards the boundary at D^s and crosses D^s to enter D_b, always following the same orientation along the x_1-axis, as $\dot{x}_1$ has the same sign on both regular domains. A continuous solution will still be obtained at $x_1 = \theta_1$ (for any α), but it could change direction along the other coordinates (see Fig. 1.2(a)). If $f_1^{D_a} - \gamma_1\theta_1$ and $f_1^{D_b} - \gamma_1\theta_1$ have opposite signs, then the two vector fields generate opposing trajectories at each point in the switching domain D^s, pushing the system in opposite ways (see Fig. 1.2(b)). A "natural solution" can then be found by letting $x(t)$ evolve on the switching plane, i.e., setting $\dot{x}_1 = 0$ for $x_1 = \theta_1$ (see [33] for more details). Note that this leads to a particular value for α given by:

$$0 = \alpha f_1^{D_a}(\theta_1) + (1 - \alpha) f_1^{D_b}(\theta_1) - \gamma_1\theta_1 \quad \Rightarrow \quad \alpha = \frac{\gamma_1\theta_1 - f_1^{D_b}(\theta_1)}{f_1^{D_a}(\theta_1) - f_1^{D_b}(\theta_1)}.$$

(The fact that $0 \leq \alpha \leq 1$ follows from the hypothesis $f_1^{D_i} - \gamma_1\theta_1 < 0 < f_1^{D_j} - \gamma_1\theta_1$, $i, j \in \{a, b\}$, $i \neq j$.) This type of solution can be "unstable" (if both vector fields point away from D^s) or "stable" (if both vector fields point towards D^s), in which case it will be called a *sliding mode solution*.

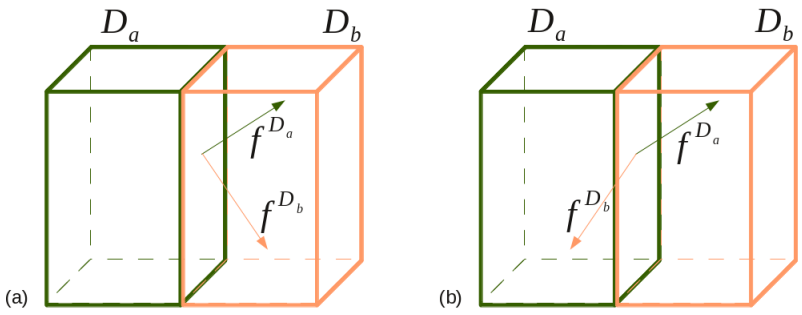

Figure 1.2: Two different configurations of vector fields on adjacent domains: (a) trajectories can cross from D_a to D_b in a natural way; (b) a solution would be to remain on the switching surface shared by D_a and D_b.

State transition diagram A trajectory will thus evolve among the regular and switching domains in the state space. Its evolution can be viewed in a more schematic form with the help of a *state transition diagram*. This form of diagram was first suggested in [29] (and then generalized to discrete models) and is composed of a set of vertexes, which represent the regular and switching domains [10], and a set of edges, which represent the possible transitions between the domains. If no sliding mode solutions are present, the switching domains can be omitted from this diagram. The regular domains can be labeled, for instance according to (1.5). The possible transitions from each regular domain depend on the location of its focal point, and hence on the parameters. To construct this diagram it is useful to use the notation (1.5) and re-write the vector field in terms of the focal points $\phi^D = (M_1^D, \cdots, M_n^D)$:

$$\dot{x}_i = \gamma_i(M_i^D - x_i), \quad \forall\, x \in D = \text{“}r_1, \ldots, r_n\text{”}.$$

Then, in general, for each i (see the example below)

$$\begin{cases} M_i^D - x_i > 0\ \forall x \in D, & \text{add arrow “}r_1 \cdots r_i \cdots r_n\text{”}\rightarrow\text{“}r_1 \cdots r_i + 1 \cdots r_n\text{”} \\ M_i^D - x_i < 0\ \forall x \in D, & \text{add arrow “}r_1 \cdots r_i \cdots r_n\text{”}\rightarrow\text{“}r_1 \cdots r_i - 1 \cdots r_n\text{”}. \end{cases}$$

If the sign of $M_i^D - x_i$ is not constant in D, then no arrow is added; if $r_i + 1 \geq \text{Max}_i$ or $r_i - 1 \leq 0$ again no arrow is added, since x_i cannot increase (or decrease) out of R. Each outgoing arrow represents the crossing of a threshold and, to maintain a more realistic description, for each arrow only one variable may cross a threshold. Hence, each vertex may have zero or up to n outgoing arrows. The resulting diagram (an example is shown in Fig. 1.3) is very useful to compactly describe and visualize the possible trajectories of the system. The qualitative behavior (oscillations, location of steady states, stability) can be inferred from the state transition diagram, but detailed knowledge is lost. For instance, for a vertex with more than one outgoing arrow, there are now several possible trajectories, but no information on the “most probable”. Likewise, it is clear that a periodic orbit of the PWA system leads to a *transition cycle* in the diagram, but the converse is not true, as a transition cycle may also correspond to a damped oscillation. To recover part of this quantitative knowledge, one possibility is to assign values or probabilities to each edge, “randomly”, or according to biological data [C23],[C13], [22], or based in the parameters of the PWA system [C12], [C4].

Example For the reduced NFκB example (1.2), the PWA system is:

$$\begin{aligned} \dot{x}_a &= k_a s^-(x_b, \theta_{ba}) - \gamma_a x_a \\ \dot{x}_b &= k_b s^+(x_r, \theta_r) - \gamma_b x_b \\ \dot{x}_r &= k_{ar} s^+(x_a, \theta_a) + k_{br} s^-(x_b, \theta_{br}) - \gamma_r x_r. \end{aligned} \tag{1.6}$$

Note that x_a and x_r have one threshold each, since they act on one variable each, while x_b has two thresholds, since activity of x_a or x_r may be triggered with different values of x_b. There is an invariant region, given by

$$\left[0, \frac{k_a}{\gamma_a}\right] \times \left[0, \frac{k_b}{\gamma_b}\right] \times \left[0, \frac{k_{ar} + k_{br}}{\gamma_r}\right]$$

since the vector field points inwards, so typically the study of the dynamics is restricted to this invariant region.

Assuming that all thresholds are inside this invariant region, and in particular satisfy the inequalities:

$$0 < \theta_a < \frac{k_a}{\gamma_a}, \quad 0 < \theta_{ba} < \theta_{br} < \frac{k_b}{\gamma_b}, \quad 0 < \frac{k_{ar}}{\gamma_r} < \theta_r < \frac{k_{br}}{\gamma_r}, \tag{1.7}$$

the state space can be partitioned into 18 rectangular regions $D = I_a \times I_b \times I_r$, where:

$$\begin{aligned} I_a &\in \left\{(0, \theta_a), \left(\theta_a, \frac{k_a}{\gamma_a}\right)\right\}, \quad I_b \in \left\{(0, \theta_{ba}), (\theta_{ba}, \theta_{br}), \left(\theta_{br}, \frac{k_b}{\gamma_b}\right)\right\} \\ I_r &\in \left\{\left(0, \frac{k_{ar}}{\gamma_r}\right), \left(\frac{k_{ar}}{\gamma_r}, \frac{k_{br}}{\gamma_r}\right), \left(\frac{k_{br}}{\gamma_r}, \frac{k_{ar} + k_{br}}{\gamma_r}\right)\right\}. \end{aligned} \tag{1.8}$$

Following (1.5), the 18 regular domains can be labeled so that (see Fig. 1.3)

$$\text{“120”} \Leftrightarrow x_a \in \left(\theta_a, \frac{k_a}{\gamma_a}\right), \; x_b \in \left(\theta_{br}, \frac{k_b}{\gamma_b}\right), \; x_r \in \left(0, \frac{k_{ar}}{\gamma_r}\right).$$

To construct the state transition diagram, the outgoing arrows from each vertex need to be computed. In domain “120” the vector field is as follows:

$$\begin{aligned} \dot{x}_a &= -\gamma_a x_a \; < 0, \\ \dot{x}_b &= -\gamma_b x_b \; < 0, \\ \dot{x}_r &= k_{ar} - \gamma_r x_r \; > 0, \end{aligned}$$

so both x_a and x_b may decrease below a respective threshold, while x_c may increase above a threshold. For this vertex, three arrows are added, one along each direction:

$$\text{“120”} \to \text{“020”}, \;\; \text{“120”} \to \text{“110”}, \;\; \text{“120”} \to \text{“121”}.$$

For system (1.6) note that, in the case $k_{br} = 0$, the system reduces to a simple negative feedback loop with three components. It has been shown in [21] that such a PWA system has a stable periodic orbit. However, for $k_{br} > 0$, a second negative loop appears which may substantially change the dynamics. In [C4] the system (1.6) was studied for sets of parameters satisfying (1.7), with the state space partitioned according to the intervals (1.8). The corresponding transition diagram is shown in Fig. 1.3, where several (five) transition cycles exist, but no sliding modes, hence only vertexes representing regular domains are depicted. In this case, the transition diagram shows that none of the regular domains contains its focal point, hence no classical equilibria exist. In fact, one can further see that the system has an unstable Filippov-type equilibrium point at $(\theta_a, \theta_{ba}, \theta_r)$, and will exhibit a periodic orbit [C4]. The periodic orbit will follow one of the five transition cycles, depending on the set of parameters.

For other regions of parameters, such as

$$0 < \theta_a < \frac{k_a}{\gamma_a}, \quad 0 < \theta_{ba} < \theta_{br} < \frac{k_b}{\gamma_b}, \quad \theta_r < \min\left\{\frac{k_{ar}}{\gamma_r}, \frac{k_{br}}{\gamma_r}\right\}, \tag{1.9}$$

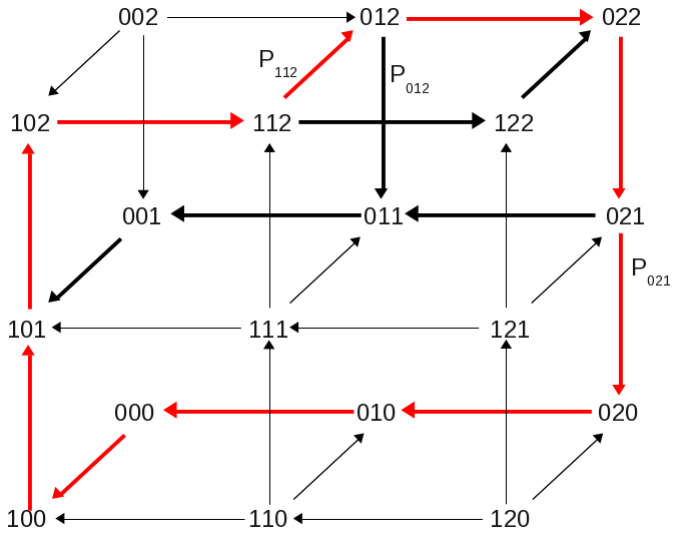

Figure 1.3: State transition diagram for system (1.6) with parameters satisfying (1.7). Each node represents one of the 18 regular domains of the state space, labeled according to (1.5). The bold arrows represent the asymptotic behavior of the system. To introduce more quantitative aspects, a *transition probability* may be assigned to the vertexes with more than one outgoing arrow. Some are indicated by P_{112}, P_{012}, and P_{021}.

it is possible to show that $(0, \theta_{br}, \theta_r)$ is in fact a stable equilibrium point (of Filippov-type, since it lies on a switching domain). To see this, observe that the set

$$S = [0, \theta_a) \times \left(\theta_{ba}, \frac{k_b}{\gamma_b}\right] \times \left[0, \frac{k_{br}}{\gamma_r}\right]$$

is invariant for the system. In S, system (1.6) with parameters (1.9) can, in fact, be reduced to a two-dimensional system,

$$\begin{aligned} \dot{x_b} &= k_b s^+(r, \theta_r) - \gamma_b b \\ \dot{x_r} &= k_{br} s^-(b, \theta_{br}) - \gamma_r r. \end{aligned}$$

since the variable x_a is strictly decreasing ($\dot{x_a} = -\gamma_a x_a$) and exerts no influence on the others. It is well known that this 2D system converges towards the steady state (θ_{br}, θ_r) [31]. Therefore, in a (sufficiently small) open neighborhood of $(0, \theta_{br}, \theta_r) \in S$, the full system (1.6) converges towards this point. In the state transition diagram, the point $(0, \theta_{br}, \theta_r)$ would be on a switching domain at the center of the rectangle formed by "011", "021", "012", and "022".

In conclusion, the parameters play an important role in determining the dynamics of the state transition diagram. Moreover, this object can be used to explore the connections between the PWA and Boolean frameworks, in order to compare and complement our knowledge on a given system.

1.4 Asynchronous Boolean networks

Boolean networks have first been introduced for modeling genetic regulatory networks by S. Kauffman and R. Thomas [61, 62]. Since then, many aspects of the Boolean formalism have been studied, such as conditions for existence of given configurations of attractors [63, 49], the role of positive or negative circuits [50, 51], or extensions to other logical frameworks, including time constraints [8, 57]. There is a natural relationship

between Boolean, discrete, and piecewise affine systems [C11], [38], which have been seen in Section 1.3 and will be further illustrated below.

Due to the rapid advances in molecular biology, regulatory networks are known to involve increasingly larger number of components, as well as greater complexity. Boolean models have thus become a most useful framework, especially in the case of large networks (on the order of 10 or more variables), with many successful applications [54, 1, 71, 11, 45, 56, 46, 22, 53, 55, 9].

Boolean networks consist of a set of nodes or variables (mRNAs, proteins or other entities) of the genetic system, and a set of logical rules, which represent the interactions between those variables (see, for instance, [69] for a review). The logical rules dictate the dynamical behavior of each variable. Boolean variables can only take two values, 0 or 1, which are often appropriate for modeling gene expression, especially when data is scarce. In this case, one can say that 0 (resp., 1) corresponds to a weak (resp., strong) expression of a gene or protein. The Boolean variables will be denoted by $X = (X_1, \ldots, X_n)$ and the state space $\Omega = \{0,1\}^n$. The time is also assumed to be discrete, $0 < t_1 < \cdots < t_k < \cdots$ and the dynamics of a Boolean model is specified by a set of logical rules, $\{F_i(X) : \Omega \to \{0,1\}, i = 1, \ldots, n\}$ which determine the state of the nodes at time t_{k+1}, given the state at time t_k: $X_i(t_{k+1}) = F_i(X(t_k))$. To simplify notation, one usually defines

$$X_i^+ := F_i(X), \quad i = 1, \ldots, n.$$

To determine the temporal evolution of the system, one must specify a mode of update, roughly, the order in which the variables progress in time. Several algorithms exist, as briefly discussed next.

Synchronous and asynchronous networks In general, there are different possible strategies to *update* the system and obtain its trajectories. One of the most common assumes that *all nodes are simultaneously updated*, that is, at each instant t_{k+1} all variables change to their new value, so

$$\forall k > 0, \ \ \forall i = 1, \ldots, n \quad X_i(t_{k+1}) = F(X_i(t_k)).$$

The corresponding networks are called *synchronous Boolean networks*.

However, the synchronous assumption is not always very realistic, as the timescales of different biological processes can vary widely (translation or transcription are generally much slower than complex binding; some timescales are given in [2]). A more general updating strategy considers that, at each instant, only one node is updated to its new value, i.e.,

$$\forall k > 0, \ \ \exists! j \in \{1, \ldots, n\} \quad X_j(t_{k+1}) = F_j(X_j(t_k)) \ \text{ and } \ X_i(t_{k+1}) = X_i(t_k)) \ i \neq j.$$

The corresponding networks are called *asynchronous Boolean networks*. There are other intermediate or mixed strategies [C23],[22], but they are typically based on these two. For all strategies, the trajectories of the Boolean network consist of a sequence of transitions among the 2^n states in Ω. There is a transition between two states $V, W \in \Omega$ if $V = X(t_k)$ and $W = X(t_{k+1})$, in which case one says that W is a *successor* of V. The set of successors of X is obtained as:

$$\sigma(X) = \{\tilde{X} \in \Omega : \ \exists j \ \tilde{X}_j = F_j(X) \ \text{ and } \ \tilde{X}_i = X_i \ i \neq j\}.$$

All possible trajectories can also be described as a directed graph with 2^n vertexes (the cardinality of Ω), with an edge connecting two vertexes whenever one state is the successor of the other. Note that there is a fundamental difference between the directed graphs corresponding to synchronous or asynchronous networks. In the former, any given state can have at most one successor, thus generating deterministic behavior, while in the latter each state can have up to n successors, generally leading to several different choices of trajectories from any given state. To understand this difference intuitively, consider that each vertex corresponds to a "large" region in the continuous state space (eg. "low" x_1 and "high" x_2); for the continuous system, each initial condition generates one trajectory, and thus several different trajectories are indeed possible from this region. The graph associated to an asynchronous network would thus capture all qualitatively distinct trajectories originating on the corresponding region in the continuous state space.

Graph theoretical representation and tools I have chosen to focus on asynchronous Boolean networks, since they permit a more realistic interpretation. The associated directed graph will be called *the asynchronous transition graph*, and its properties can be studied using graph theoretical tools. A directed graph can be decomposed into *strongly connected components* (SCCs), which are maximal subsets of vertexes where every pair is mutually reachable (two vertexes are said to be mutually reachable if there are directed paths linking one vertex to the other). SCCs can contain a single state or several states (the whole state space can constitute a single SCC, for some particular models). Hence, SCCs can have outgoing paths directed towards (states contained in) other SCCs. An SCC that contains no outgoing path is called a *terminal SCC* or *attractor*, since any trajectory that reaches it cannot leave. The SCCs of a graph can in fact be viewed as the vertices of a new acyclic graph, where the SCCs are hierarchically organized into levels such that, from each level, there can only be transitions to a higher level. This analysis facilitates the identification of the attractors, which contain the asymptotic behavior of the system.

Moreover, the asynchronous transition graph can also be described as a Markov chain (represented by a $2^n \times 2^n$ matrix $P = [p_{ij}]$), if a probability of transition is associated to each edge in the graph. Thus p_{ij} denotes the probability that there exists a transition from state i to state j ($p_{ij} = 0$ if no such edge exists). In the most simple case, the probabilities can be equally assigned among the outgoing edges of a vertex:

$$p_{ij} = \frac{1}{\#\text{successors of state } i}.$$

More generally, different weights may be assigned to different edges according to biological knowledge, for instance, due to the frequency or timescale of the process described by that edge (see, for instance, the approaches in [C23] or [C13]). For instance, in Fig. 1.3, there are two possible pathways from state "112": a probability P_{012} has been assigned to the edge $112 \to 012$; then the edge $112 \to 122$ will have a probability $1 - P_{012}$. Similarly for the states "012" and "021". Several questions, concerning transient behavior or reachability from a given state, can also be answered from analysis of the transition graph. Indeed, the transition matrix P contains useful quantitative information on a qualitative system: the ij-th entry, p_{ij}^k, of the k-th power of the matrix P denotes the probability that a trajectory follows a path from vertex i to vertex j in k steps. Therefore, j is reachable from i if, for some $k \geq 1$, $p_{ij}^k > 0$. If i is a single state attractor, then $p_{ii} = 1$ and $p_{ij} = 0$ for all $j \neq i$. If a set of states $\mathcal{I} = \{i_1, \ldots, i_L\}$ constitutes an attractor of the system, then for any $i_\ell \in \mathcal{I}$, $p_{i_\ell j} > 0$ if and only if $j \in \mathcal{I}$. By separating the rows and columns relative to the attractors, the matrix P can be written as [25]

$$P = \begin{pmatrix} P_a & 0 \\ R_a & \tilde{P} \end{pmatrix}$$

where P_a is of dimension $2^L \times 2^L$ and $\tilde{P}$ is $2^{N-L} \times 2^{N-L}$ has no attractors, so there is a path leading from every state in $\{N-L, \ldots, N\}$ to some state in $\{1, \ldots, L\}$. Then, $(I - \tilde{P})$ is invertible and

$$\begin{pmatrix} T_1 \\ \vdots \\ T_{N-L} \end{pmatrix} = (I - \tilde{P})^{-1} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

provides the *estimated time for absorption* of the state $i \in \{N-L, \ldots, N\}$ into an attractor (if the probabilities are normalized, then T_i gives the number of steps needed on average to reach an attractor from state i).

As will be seen in the example below, a Boolean model can be derived from a piecewise affine system. In the case, the probabilities p_{ij} can be assigned based on the parameters of the PWA system, as in [C12], [C4], for instance.

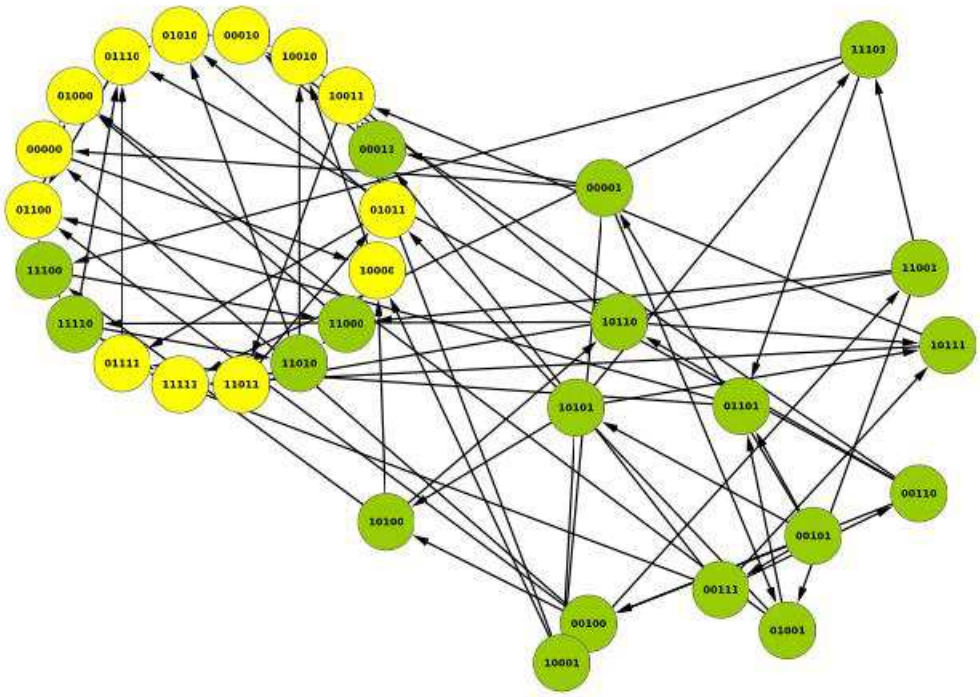

Figure 1.4: The asynchronous transition graph for the Boolean network (1.10). The 18 vertexes organized in a circle correspond to the permissible states. Within these, the light colored vertexes compose the only attractor (terminal strongly connected component) of the network (see also Fig. 1.5).(Figure drawn using the open source software platform Cytoscape [58].)

Example To write the NFκB example above as a Boolean model, the methodology proposed in [C11] can be used to extend the discrete system generated by the PWA model. Briefly, for each positive activity level, one Boolean variable is added to the system, so that a discrete variable x with levels $0, 1, \ldots, d$ is described by d Boolean variables X_1 to X_d, which satisfy: $X_j \geq X_{j+1}$. In other words, if $x = 2$ then $X_1 = X_2 = 1$, while the state $X_1 = 0$ and $X_2 = 1$ is biologically meaningless (since a very high expression level naturally entrains all the intermediate levels). The extended Boolean state space can thus be divided into "forbidden" and "permissible" states, where the former can have any arbitrary dynamical evolution and the latter must represent the dynamics of the discrete system generated by the PWA model. The logical rules for the extended model should thus mimic the PWA model (for the permissible states) and also guarantee that no transitions are allowed from the permissible to the forbidden states (for a biological meaningful model). In [C11], a method is proposed to generate appropriate Boolean variables and assign logical rules that satisfy these consistency rules. The Boolean model corresponding to (1.6) has one variable to describe species x_a (A) and two variables each to describe x_b (B_1, B_2) and x_r (R_1, R_2). The model is then:

$$\begin{aligned} A^+ &= \neg B_1 \\ B_1^+ &= B_2 \vee R_2 \\ B_2^+ &= B_1 \wedge R_2 \\ R_1^+ &= A \vee R_2 \vee (\neg B_2 \wedge R_1) \\ R_2^+ &= A \wedge R_1. \end{aligned} \tag{1.10}$$

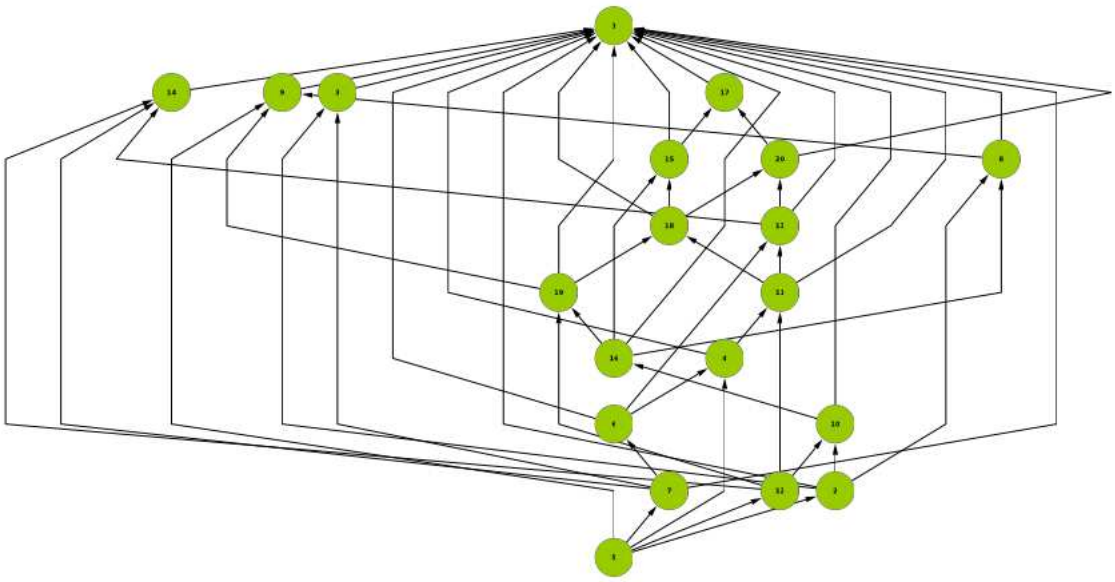

Figure 1.5: The hierarchical decomposition into strongly connected components of the graph of Fig. 1.4. The top vertex corresponds to the only attractor of the network, which contains the states corresponding to the light colored vertexes of Fig. 1.4. All other vertexes correspond to SCCs which are composed of a single state.(Figure drawn using the open source software platform Cytoscape [58].)

To construct the asynchronous transition graph, one proceeds as follows: (1) for each state $X \in \Omega$, compute the possible variable changes from the synchronous Boolean table; (2) then consider only one change at a time, to obtain all the successors of X ($Y_1, \ldots, Y_\ell$) and draw an edge from X to each Y_i. For example:

$$X = (1, 1, 1, 0, 0) \quad \Rightarrow \quad X^+ = (0, 1, 0, 1, 0)$$

so the possible successors are

$$(0, 1, 1, 0, 0), \quad (1, 1, 0, 0, 0), \quad (1, 1, 1, 1, 0)$$

by allowing only the variable, respectively, A, B_2, or R_1 to change at each time. It is not difficult to check that there are indeed no transitions from the "permissible" to the "forbidden" states, as can also be seen by observation of the corresponding asynchronous transition graph, in Fig. 1.4, where the 18 vertexes organized in a circle correspond to the permissible states. Note that these 18 vertexes correspond to the states in the state transition diagram of the PWA system shown in Fig. 1.3.

The graph of Fig. 1.4 can be decomposed into SCCs which shows that there is a single attractor, labeled "1" (see Fig. 1.5). Furthermore, the Boolean states contained in this attractor are those which are shown in light color in Fig. 1.4 and correspond to the asymptotic states (those connected by bold arrows) in the state transition diagram of the PWA system (Fig. 1.3).

In conclusion, there is indeed a clear correspondence between the state transition diagram of the "hybrid" PWA model (1.6) (with parameters satisfying (1.7)) and the asynchronous transition graph the Boolean system (1.10); both frameworks capture the (same) *qualitative asymptotic behavior* of the biological system, although the "hybrid" approach further provides continuous solutions across the regular domains.

Boolean networks with inputs and outputs To study other properties, such as the interconnection of two networks [C32],[C3], it is useful to introduce Boolean models into a classical control theoretical framework [59].

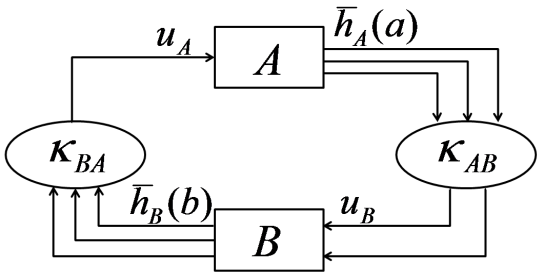

Figure 1.6: The interconnection of two multiple-input multiple-output systems A and B.

A *Boolean multiple input-multiple output (MIMO) system* is characterized by its state space Ω, a set of p inputs $u \in \mathcal{U} = \{0,1\}^p$, a set of q outputs given by an *output function* $h : \Omega \to \mathcal{H}$, with $\mathcal{H} = \{0,1\}^q$, and a logical vector function $F : \Omega \times \mathcal{U} \to \mathcal{H}$.

The inputs typically represent quantities that can be regulated or controlled by a scientist, while the output functions represent quantities that can be measured. In the Example, the input might be a stimulating substance such as Tumor Necrosis Factor (TNF) which is known to affect the NFκB network [47] and the output function might be, for instance, the expression of IκB:

$$u = \text{TNF}, \quad h(X) = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}.$$

At a basic level, TNF activates IKK and will thus have a negative effect on IκB, hence one way to include the effect of input u on (1.10) is as follows:

$$\begin{aligned} B_1^+ &= B_2 \vee R_2 \\ B_2^+ &= \neg u \vee (B_1 \wedge R_2). \end{aligned}$$

For each fixed u, the Boolean MIMO system has a specific asynchronous transition graph, $G = G(u)$. In this case, the graph $G(0)$ coincides exactly with be the one represented in Fig. 1.4. The case $u = 1$, would imply $B_2 \equiv B_1 \equiv 0$ and the resulting changes.

Given two Boolean MIMO systems, their interconnection can be described by two feedback functions that transform the outputs of one system into the inputs of the other. Let the two systems, A and B, be characterized by $(\Omega^A, \mathcal{U}^A, \mathcal{H}^A)$ and $(\Omega^B, \mathcal{U}^B, \mathcal{H}^B)$, with output functions $\bar{h}_A$ and $\bar{h}_B$ and logical rules F^A, F^B. The interconnection of A and B can be described by two feedback functions that transform the outputs of one system into the inputs of the other:

$$\kappa_{AB} : \mathcal{H}^A \to \mathcal{U}^B, \quad \kappa_{BA} : \mathcal{H}^B \to \mathcal{U}^A.$$

Consider the composition of κ_* and $\bar{h}_*$: $h_A(a) = \kappa_{AB}(\bar{h}_A(a))$ and $h_B(b) = \kappa_{BA}(\bar{h}_B(b))$. Then, the *interconnection of A and B* is the Boolean system Σ, with no inputs or outputs, with state space $\Omega = \Omega^A \times \Omega^B$, and Boolean rules $F^\Sigma : \Omega \to \Omega$ constructed in the following way:

$$F^\Sigma(a,b) = \left(F^A(a; h_B(b)), F^B(b; h_A(a))\right). \tag{1.11}$$

Let $G^A(u)$ (resp., $G^B(v)$) denote the asynchronous transition graph of system A under input u (resp., system B under input v). Define $\sigma_{A,u}(a)$ to be the set of successors of state a for system A under input u, that is, in the asynchronous transition graph $G^A(u)$ (similarly for $\sigma_{B,v}(b)$). The *successors* of an element of Ω are again computed according to the asynchronous updating strategy, and they are of the form

$$\left\{(a,\tilde{b}), (\tilde{a},b) \in \Omega : \ \ \tilde{a} \in \sigma_{A,h_B(b)}(a) \text{ and } \tilde{b} \in \sigma_{B,h_A(a)}(b)\right\}. \tag{1.12}$$

Intuitively, trajectories of the interconnection system will evolve either in $G^A(h_B(b))$ with part b fixed, or in $G^B(h_A(a))$ with part a fixed. At any update, the trajectory can "switch" between graphs, depending on the successors at that instant. More details asnd results are left to Chapter 5, which discusses this recent work.

1.5 Outline

The next chapters collect a selection of my published articles, ranging from 2005 to 2013, chosen to showcase my work in the analysis of dynamical systems applied to signaling and genetic regulatory networks. By putting together this collection, I wished to highlight the evolution and richness of my theoretical work as inspired by questions and problems from biological regulatory networks. This *mémoire*, as a collection of articles, is divided into four chapters, each dedicated to a different problem, as summarized next.

Chapter 2 is composed of two articles dealing with introductory questions on how to model the various biological events in a network and compare the different formalisms. The first article is a short introduction to modeling and analysis of genetic regulatory networks, using continuous and piecewise affine models, which is published in a book based on a series of courses taught in the scope of the Master of Science program on Computational Biology and Biomedicine at the Université Nice Sophia Antipolis, France. The second article is on the mathematical comparison between Boolean, multi-level discrete and piecewise affine models.

Chapter 3 collects four articles dedicated to the quantitative analysis of various dynamical properties: robustness with respect to changes in the parameters and convergence to steady states, dependence of the system on its different timescales, as well as probabilistic approaches for predicting which steady state or periodic will more likely be attained. The papers also show how to combine Boolean and piecewise affine networks, to obtain more realistic models and quantitative analysis. These four papers deal with different biological examples, including the segment polarity network of the fruit fly *Drosophila melanogaster* and a mammalian apoptosis network.

Chapter 4 comprises two articles on qualitative techniques for control of piecewise affine systems in the plane. The first article uses the rectangular partition of the state space to consider a problem where the available measurements consist of an interval for each concentration, and the possible actions (or control inputs) on the system are constant in each rectangular region and would correspond to switching gene expression on or off in each region. The second article studies a system where the state space is partitioned into conical regions, whose order and respective focal points change with the regulatory function (i.e., activation, inhibition) of one of the variables on the others. The various dynamical behaviors induced by this global control are fully characterized.

Chapter 5 consists of two articles that describe and prove a theoretical method for the analysis of large Boolean networks, by the interconnection of two smaller modules. The asymptotic behavior of the full network obtains by studying the two smaller modules, thereby greatly reducing the computational effort and time.

One final chapter summarizes the main results contained in this *mémoire*, and suggests further applications of the methods described here, as well as future research directions.

The Appendices contain further information, such as my Curriculum Vitae, complete list of publications, and the abstract of my PhD. thesis. The original publications of the ten articles mentioned in the four chapters are included in Appendix B.

Chapter 2

Modeling genetic regulatory systems: from continuous to Boolean networks

In general, for any mathematical framework, there are several possible ways of representing a given biological process, so the construction of the model requires an inspection of the network of interactions. An introduction to modeling genetic regulatory networks is given in §2.1, using continuous, piecewise affine and discrete models.[1]

The introductory paper §2.1 is intended for graduate students with either mathematical or biological backgrounds and gives a detailed description of how to model activation, transcription and translation events, and then to assemble the various processes in a single model and analyze the dynamics of this model. For continuous systems, some simplifying hypotheses and their validity are discussed, namely the quasi-steady state approximation which is based on the fact that different processes have different timescales, implying that the model can be reduced under appropriate conditions. Classical methods are recalled for continuous ordinary differential equations, such as verification of positivity, computation of the steady states of a model, their local analysis by linearization around the steady states and inspection of the eigenvalues. Not so classical methods are also recalled, such as Tikhonov's Theorem for simplifying a model based on different timescales' arguments. For piecewise affine systems, it is recalled that solutions can be defined according to a construction due to Filippov, by interpreting the system of equations as a differential inclusion (see also above [...]). A basic example - the bistable switch - is used to illustrate model assembly and analysis, both in the continuous and the piecewise affine frameworks.

In view of all the different formalisms available, it becomes important to have an idea of whether and how they can be compared. Furthermore, if one of the goals is to "transfer" or exchange information between model formulations, then one would like to guarantee that the different formulations do model the same system, and that parameters (whenever appropriate) can be related. The article §2.2 presents some methodologies to compare and interchange models in the piecewise affine, discrete and Boolean formalism.

As can be seen in the article §2.1, there is a natural way to relate continuous and piecewise affine models if the former use saturated, sigmoid type activity functions. Letting the cooperativity exponent tend to infinity, the sigmoid function converges to a step-like, Heaviside function, whose point of discontinuity (or activity threshold) is given by the half-maximal concentration. The continuous set of differential equations is thus transformed into a set of piecewise affine differential equations, where the vector field has a finite number of discontinuities, at the threshold points of the step functions.

A piecewise affine system can itself be straightforwardly related to a multi-valued discrete system, given the

[1] This article is the second chapter of a book based on a series of courses taught in the scope of the Master of Science program on Computational Biology and Biomedicine at the Université Nice Sophia Antipolis, France. Since 2009, I have regularly taught in this program, on modeling and analysis of gene regulatory networks using continuous and piecewise affine models. In this article, together with Jean-Luc Gouzé, I have been responsible for section 2.

partition of the state space into hyper-rectangles (see above [...]). Indeed, recall that an hyper-rectangle is defined by a product of intervals in each variable, defined by the thresholds, such as $[\theta_1^{i_1}, \theta_1^{i_1+1}] \times \cdots \times [\theta_n^{i_n}, \theta_n^{i_n+1}]$. Now, if a (continuous) variable has d_L thresholds, one can postulate that the corresponding discrete variable has d_L states, so that each hyper-rectangle corresponds to a single state of the discrete model, for instance $(i_1, \ldots, i_n)$. The dynamics of the discrete model can then be established by looking at the possible transitions between hyper-rectangles to construct a transition graph between discrete states. To be more realistic, only transitions between adjacent hyper-rectangles (i.e., that share at least one face) are allowed, corresponding to crossing only one threshold at a time.

Finally, several methods can be devised for relating discrete and Boolean models. In §2.2, an idea due to [65] is used to obtain a Boolean model from a multi-valued discrete model by extending the number of variables: if a discrete variable X has d_L states, than $d_L - 1$ Boolean variables are created, X_1, X_{d_L-1} to describe X. The Boolean variable X_j is at 1 if and only if the discrete variable X is in state j or higher. However, this idea should to be analyzed with care, since it generates non-realistic (which will be called "forbidden") states (if $X_j = 1$ then it makes no sense to have $X_k = 0$ for $k < j$). Thus, in the Boolean transition graph, one wishes to avoid transitions from the permissible to the forbidden states. We have introduced and characterized a map that transforms discrete into Boolean models and vice-versa, while preserving the required biological properties.

Under these conditions, in the transformation from continuous to discrete models, all behavior on threshold regions is lost, such as sliding mode solutions. As illustrated by an example, a sliding solution will be replaced by two arrows with opposite orientations linking the same pair of states. In this case, an alternative representation can be suggested, by introducing an intermediate discrete state that has two incoming arrows, on from each of the other two states.

2.1 Modeling and analysis of gene regulatory networks, by G. Bernot, J.-P. Comet, A. Richard, M. Chaves, J.-L. Gouzé and F. Dayan. In "Modeling in Computational Biology and Biomedicine", F. Cazals and P. Kornprobst Eds, Springer-Verlag Heidelberg (2013), pp. 47-80.

Article by G. Bernot, J.-P. Comet, A. Richard, M. Chaves, J.-L. Gouzé and F. Dayan. In "Modeling in Computational Biology and Biomedicine", F. Cazals and P. Kornprobst Eds, Springer-Verlag Heidelberg (2013), pp. 47-80.

2.2 Comparison between Boolean and piecewise affine differential models for genetic networks, by M. Chaves, L. Tournier and J.-L. Gouzé. Acta Biotheoretica, 58(2)(2010), pp. 217-232

Article by M. Chaves, L. Tournier and J.-L. Gouzé. Acta Biotheoretica, 58(2)(2010), pp. 217-232.

Chapter 3

Quantitative methods: analysis of dynamical properties

This chapter presents some methods adapted to study the main properties of a biological system, while trying to use only a reduced family of parameters. These methods are designed to make the most of the available data with a minimum amount of mathematical machinery. Although apparently very schematic and qualitative, these models and methods yield quantitative results used to make verifiable predictions, discriminate between different modeling hypotheses, or predict the most likely outcome of a trajectory.

The four articles included here also illustrate the evolution of the methodology to incorporate more quantitative aspects into a discrete framework, while maintaining the easier analytical tractability and intuition provided by Boolean or piecewise affine models.

The first article §3.1 can in fact be said to mark the beginning of my work on Boolean networks. It proposes three asynchronous updating algorithms that incorporate timescales and some parameters back into Boolean models. As recalled in the Introduction (Section 1.4), an underlying assumption for synchronous Boolean models is that all the processes described have approximately the same timescales, and so their states evolve and change (almost) simultaneously. However, it is well known that not all biological processes happen at the same timescale (for instance, binding of proteins is much faster than transcription or translation). In §3.1 we thus sought to introduce a more realistic time dynamics. Namely, our most general asynchronous algorithm considers that the next node to be updated is chosen randomly; the random order algorithm requires each node to be updated exactly once in each time interval (but following a random order at each round of updates); the third (separation of timescales) algorithm also stipulates a full round of updates but, in each round, the order is chosen according to the timescales (for instance, updating first all the protein nodes are then all the mRNA nodes). All these are illustrated by application to the *Drosophila melanogaster* segment polarity Boolean network developed by R. Albert and H. Othmer [1], to infer its robustness properties with respect to timescales of the biological processes. This model describes pattern formation between stages 8 and 11 of embryonic development and has six distinct steady states. One of these corresponds to the wild type (the "normal" phenotype) and two others correspond to some mutant phenotypes; the remaining steady states are variations on the wild type. Fixing the initial condition (corresponding to stage 8 of development), the dynamical evolution of this Boolean model with the asynchronous or random order algorithms leads most often, but with a relatively low frequency (around 55%), to the wild type steady state. In contrast, the two-timescale algorithm has a frequency of 87.5%, a result that can be explicitly calculated based the order of nodes updates. Furthermore, this algorithm can also be characterized by a Markov chain with two absorbing states, the wild type and one of the mutant phenotypes. Overall the results indicate that the network has a high sensitivity to timescales, but is nevertheless robust to variability once the timescales are fixed (i.e., within a class of nodes, the order may vary). Further questions raised by the Drosophila network were also explored in [C21] (combining Glass-type

models with Boolean activities; see also §3.2) or [C18] (studying the robustness of the network of interactions under cell division).

The second article (§3.2) develops a Boolean model for an apoptosis pathway and its interaction with the transcription factor NFκB (Nuclear Factor κ) network, based on continuous models for each of these modules [17, 47]. The dynamics of the system is studied using Glass-type models (see Section 1.3 in the Introduction), as had been proposed in [C18]. We conclude that there are different possible dynamics: oscillatory behavior in the presence of external stimulation, or convergence to either one of two steady states, which represent apoptosis (cell death) or a living cell, in the absence of stimulus. Several quantitative aspects could be reproduced and compared to experimental data in the literature, such as the average interval between two peaks for the oscillatory solutions. The model was used to test three different hypotheses on the interconnection between the NFκB and the apoptosis pathways, which is not clear from experimental data. Based on Monte Carlo tests and comparison to biological observations, at least one of the hypotheses could be discarded, and the most likely model could be chosen. This Boolean model was posteriorly further analyzed in [C13], where the subsystems responsible for each of the asymptotic behaviors (bistability or oscillations) were identified.

The last two articles in this chapter focus on methods for analysis of piecewise affine systems. The article in §3.3 reviews the main properties of this class of systems and proposes a method to relate the transition graph to the parameters of the model (activity thresholds, synthesis and degradation rates). In general, there may be transitions from one hyper-rectangle to several of its adjacent neighbors, and the transition graph contains no information on the frequency of each transition. Uniqueness of solutions in each hyper-rectangle implies that there will be well defined regions where every initial condition leads to a trajectory that evolves to the same adjacent hyper-rectangle. Some suggestions can be found in the literature on how to assign a probability to each edge on the transition graph, based more or less on biological arguments. To enforce the relation with the parameters of the PWA system, our idea in §3.3 is to let the probability of transition be proportional to the area of the region of initial conditions that evolve across that edge. Since solutions of the PWA are easily written inside each rectangle, the area of this region can also be analytically computed in terms of the parameters. This task becomes straightforward in the case of at most two transitions from each hyperrectangle. This method can also be used to estimate (some of the) parameters of the PWA system, for instance, by repeating the same experience with initial conditions in a given hyperrectangle, and counting the frequency of the outcomes.

The article in §3.4 extends this definition to consider the "memory" of the system, that is, the probability of transition depends both on the current and the previous hyper-rectangle crossed by the trajectory. We observe that the previous history of the trajectory prevents some transitions to happen, thus refining the procedure. This method is applied to the analysis of a system composed of two intertwined negative loops (see the example given in the Introduction). The transition graph shows that there are five distinct possible periodic orbits. Using our definitions of probability of transition, given any set of parameters, we can predict which of the periodic orbits will most likely be reached by the PWA system. Numerical simulations show that the method using the one-step probability definition correctly predicts the orbit on about 60% of the cases, while the two-step definition is correct on 70%. Thus, these two articles provide an approximate method to find whether a system contains a given periodic orbit, as well as indications on how to modify (or control in some sense) the parameters to drive the system to a desired orbit.

3.1 Robustness and fragility of Boolean models for genetic regulatory networks, by M. Chaves, R. Albert and E.D. Sontag. J. Theoretical Biology, 235(3)(2005), pp. 431-449

Article by M. Chaves, R. Albert and E.D. Sontag. J. Theoretical Biology, 235(3)(2005), pp. 431-449.

3.2 Regulation of apoptosis via the NFkB pathway: modeling and analysis, by M. Chaves, T. Eissing and F. Allgöwer. In "Dynamics on and of complex networks: applications to biology, computer science and the social sciences", N. Ganguly, A. Deutsch and A. Mukherjee Eds, Birkhauser Boston, 2009, pp. 19-34

Article by M. Chaves, T. Eissing and F. Allgöwer. In "Dynamics on and of complex networks: applications to biology, computer science and the social sciences", N. Ganguly, A. Deutsch and A. Mukherjee Eds, Birkhauser Boston, 2009, pp. 19-34.

3.3 Piecewise affine models of regulatory genetic networks: review and probabilistic interpretation, by M. Chaves and J.-L. Gouzé. In "Advances in the Theory of Control, Signals and Systems, with Physical Modelling", J. Lévine and P. Müllhaupt Eds, Springer-Verlag Heidelberg, LNCIS 407(2010), pp.241-253

Article by M. Chaves and J.-L. Gouzé. In "Advances in the Theory of Control, Signals and Systems, with Physical Modelling", J. Lévine and P. Müllhaupt Eds, Springer-Verlag Heidelberg, LNCIS 407(2010), pp.241-253.

3.4 Probabilistic approach for predicting periodic orbits in piecewise affine differential models, by M. Chaves, E. Farcot and J.-L. Gouzé. , Bull. Math. Biology, 75(6), pp. 967-987,2013

Article by M. Chaves, E. Farcot and J.-L. Gouzé. , Bull. Math. Biology, 75(6), pp. 967-987,2013.

Chapter 4

Qualitative methods: design of control strategies

One of the applications of model construction and analysis is the investigation of strategies to control or regulate the true system. For a genetic regulatory system, the "controls," or actions that can be performed on the system, amount to experimental techniques such as adding activators/inhibitors, over-expression or knock-out of a gene, or inserting an inducer-controlled plasmid to externally regulate the transcription of a given gene. The outputs are the measurements on the system which may include concentrations, reaction rates, relative increases in the expression of certain genes, and so on.

As discussed in the Introduction, both inputs and outputs depend on the experimental techniques and most of them are still of a qualitative form. The data available are becoming more smooth and detailed, with high frequency sampling, but many outputs can be viewed as "gene or protein X is strongly (or weakly) expressed", which can be readily interpreted in a piecewise affine model as "X is above (or below) the threshold θ". The inputs are usually more difficult to model: for plasmid constructions, addition of a certain amount of inducer leads to some "m-fold increase" in the transcription of a gene, but it is not always clear how to quantify an over-expression or knock-out.

Classical control techniques require inputs and outputs to be quantitative and continuous functions of time, with small errors, which is not the case for biological regulatory networks. The article in §4.1 suggests an alternative control strategy, based on qualitative inputs and outputs. It is assumed that the production rate of one mRNA or protein is controllable (through a plasmid, for instance), with the following actions: the "do nothing" or nominal behavior with an appropriate amount of inducer, turning gene expression ON (resp., OFF) by adding a sufficiently large (resp., small) amount of inducer. The input u will be, respectively, 1, very large or very small. For the outputs, it is assumed that only the region of the state space where the trajectory evolves at the current instant is known, that is, at each instant, only an interval for each variable is known ($[\theta_1^{i_1}, \theta_1^{i_1+1}] \times \cdots \times [\theta_n^{i_n}, \theta_n^{i_n+1}]$). Under these conditions, our results show that it is still possible to exactly lead a system to a desired steady state or even to remain close to a periodic orbit. This illustrates the power of the PWA framework to help controlling a system under limited data.

Further work on control strategies using only qualitative data is being developed in the context of the ANR project GeMCo, where the goal is to control the growth rate of bacteria by implementing a new genetic circuit (see also the Discussion). Along similar lines, in [C33] and [C31] we have also considered the possibility of designing observers for PWA with qualitative inputs and outputs. The results show that, for a simple example, it is indeed possible to construct an observer that approaches the trajectory of the system either in finite time, or asymptotically with a user-regulated velocity, depending on the initial conditions of the observer.

On a different perspective, the other article in this chapter (§4.2) studies the qualitative dynamics of a genetic/metabolic system induced by different forms of genetic regulation. The system is composed of a

metabolic chain where each flux is regulated by an enzyme and, in their turn, the genes coding for the enzymes are regulated by one of the metabolites. After model reduction due to different timescales, this system can be represented by a PWA system whose state space is partitioned into cones. Different configurations of the cones and their focal points are possible, depending on the genes' positive or negative regulation by the metabolite, and each configuration leads to its own dynamics. Our analysis suggests that the interconnection of metabolic and genetic networks can exhibit a wide range of behaviors, from mono- and multi- stability to periodic oscillations, as seen by our complete classification of the system according to its configurations. It is interesting to note that oscillatory behavior can happen only in the case of operon-like regulation, that is a family of genes that are all similarly regulated, as in the case of the *lac* operon in *Escherichia coli*. In contrast, for individual gene regulation, we observe that a fairly simple regulatory structure can lead to more robust, globally stable behavior or more versatile multistable designs.

4.1 Exact control of genetic networks in a qualitative framework: the bistable switch example, by M. Chaves and J.-L. Gouzé. Automatica, 47(2011), pp. 1105-1112

Article by M. Chaves and J.-L. Gouzé. Automatica, 47(2011), pp. 1105-1112.

4.2 Multistability and oscillations in genetic control of metabolism, by D.A. Oyarzún, M. Chaves, and M. Hoff-Hoffmeyer-Zlotnik. J. Theoretical Biology, 295(2012), pp. 139-153

Article by D.A. Oyarzún, M. Chaves, and M. Hoff-Hoffmeyer-Zlotnik. J. Theoretical Biology, 295(2012), pp. 139-153.

Chapter 5

Network interconnections: transient and asymptotic dynamics

This chapter is dedicated to the problem of analyzing "large" networks of interactions. Models of biological systems very often comprise a large number of species, and can easily have more than 15/20 components (see, for instance, the *Drosophila* segment polarity network or the apoptosis network studied in Chapter 3). In a continuous (or also a piecewise affine) framework, it is in general not possible to analytically study models with more than three differential equations, unless the mathematical system belongs to a class with special properties (see, for instance, my previous work on zero-deficiency biochemical networks, with polynomial vector fields [C27], [C26], [C22], or other classes of networks [35, 5]). The study of large models with continuous frameworks is thus usually limited to computer simulations, which provide only a partial overview of the models' dynamical properties.

As mentioned in the Introduction, the Boolean formalism allows a theoretical study of the dynamics of large models but, nevertheless, some of the graph theoretical tools become very heavy to implement in a reasonable time (such as the decomposition of the state transition graph into strongly connected components, whose complexity increases with 2^n, where n is the number of variables in the network).

At the same time, many systems are obtained by assembling or interconnecting two or more, smaller modules. This is indeed the case for the *Drosophila* network cited above, which is formed by interconnecting four "cells", that is four similar models describing the network of interactions in a single cell.

Thus, in the articles §5.1 and §5.2, we propose a novel idea to study the dynamical properties of a large network as the interconnection of two smaller modules. An *interconnection* of two systems, both having a set of inputs and outputs, is here defined in the classical control theoretical sense: the coupling between the two modules is represented by a pair of functions that describe how the outputs of one system are consistently transformed into the inputs of the other. A new object, called the *asymptotic graph* (G^{as}), is introduced in §5.1, which requires only the knowledge of the state transition graphs of the two modules, as well as the two coupling functions. The nodes of the asymptotic graph are, roughly, the products of the *semi-attractors* (members of an attractor which share the same output) of the two modules. The transitions between nodes are obtained by supposing that the dynamics of each module is faster than the coupling dynamics. It is shown that the set of attractors of the asymptotic graph contains a representative of each of the attractors of the interconnected system. The asymptotic graph can, however, contain more attractors than those of the interconnected system since, due to its construction, it may not contain information relative to all possible paths in the transition graph of the interconnection. Thus, this method is guaranteed to *recover all the attractors of the large network*, but may also generate some spurious attractors. Some partial results are given to help decide whether an attractor is spurious or not.

To further understand the generation of spurious attractors, a more comprehensive state transition graph

called the *cross-graph* is constructed in §5.2. Its nodes are the products of the *semi-SCCs* (members of an SCC which share the same output) of the two modules, and the edges are obtained from the two module's asynchronous transition graphs. It is shown that there is a one-to-one correspondence between the attractors of the cross-graph and those of the interconnected system. Intuitively, the cross-graph reproduces all the paths of the interconnected system but, for that same reason, it may be as heavy to compute as the transition graph of the interconnection itself (think of the limiting case where each SCC is a single state). Therefore the cross-graph is interesting from an analytical point of view, as the complement of the asymptotic graph, but not as a computational method.

Two directions for application of the asymptotic graph have been illustrated in §5.2. The first is a direct application to find the attractors of a composed system from two well studied models (as in the *Drosophila* segment polarity network). The second requires the prior decomposition of the large network into two smaller modules (as in the apoptosis model). The latter direction opens up a whole new family of questions relative to network decomposition (see also the Discussion below). There are a few classical graph partitioning algorithms, such as spectral techniques or hierarchical clustering algorithms, but these are not always appropriate as most of them aim to minimize the number of edges between modules. Among other observations, for a most successful application of the asymptotic graph, the two modules should be approximately of the same size, and contain the least number of inputs/outputs possible.

5.1 Predicting the asymptotic dynamics of large biological networks by interconnections of Boolean modules, by M. Chaves and L. Tournier. Proc. 50th IEEE Conf. Decision and Control and European Control Conf., Orlando, USA, Dec. 2011

Article by M. Chaves and L. Tournier. Proc. 50th IEEE Conf. Decision and Control and European Control Conf., Orlando, USA, Dec. 2011.

5.2 Interconnection of asynchronous Boolean networks, asymptotic and transient dynamics, by L. Tournier and M. Chaves. Automatica, 49:884-893, 2013

Article by L. Tournier and M. Chaves. Automatica, 49:884-893, 2013.

Chapter 6

Discussion and perspectives

Throughout this collection of articles a research direction emerges aimed at the development of predictive mathematical methods for the analysis of biological regulatory networks. The work presented here indicates a clear set of guidelines that lead to informative results on the dynamical properties of the system. The usefulness of abstract formalisms was illustrated by their relevance to deal with the currently available data and experimental techniques, their suitability for rigorous mathematical analysis, and their capacity for finally generating quantitative and not so abstract results.

6.1 Problems motivated by biological regulatory networks

A family of mathematical questions and problems specifically inspired by biological regulatory networks have been studied in these papers. They also cover a variety of examples from generic systems such as the bistable switch to signaling (apoptosis network) and genetic networks (*Drosophila* segment polarity or *E. coli* genetic machinery).

One of the problems is that of understanding cell-to-cell variability and its influence on the robustness of the system with respect to environmental perturbations, for instance checking the stability or not of steady states (corresponding to phenotypes) under large or small parameter variability (Sections 3.1, 3.2). Similarly, the robustness of the network to changes in the timescales of the various biological processes has also been studied (Section 3.1). Another problem is model discrimination: in the presence of incomplete knowledge, it is useful to test different network interconnections and compare the results against known data (Section 3.2).

Parameter estimation is common to all physical systems and was also briefly considered here (Sections 3.3, 3.4), where the difficulty lies essentially on the type of data and the experimental techniques available.

The design and construction of synthetic biological regulatory networks has also motivated various problems, such as finding strategies to control/regulate the system towards a desired steady state or dynamic behavior, which was explored in Chapter 4.

The analysis of large networks is another frequent problem in biology. The method proposed here was developed for Boolean networks and views the large network as an interconnection of two smaller networks. The method applies control and graph theoretical tools to obtain the global (asymptotic) behavior of the interconnected system, using only the properties of the two smaller modules (Chapter 5).

6.2 Combining PWA and Boolean models

Both formalisms are well adapted to describe the dynamics of signaling and genetic regulatory networks [30, 62]. Boolean models require no parameters but provide qualitative information based only on the network of

interactions and their logic. PWA models provide a continuous but still abstract description of the systems, and require a reduced family of parameters: activity thresholds, synthesis and degradation rates.

Boolean models are most useful for modeling large networks (let's say with more than five components), for which they provide a good intuition on the possible trajectories and asymptotic behavior of the system. However, for small networks the results may be less informative. In contrast, PWA models are most useful for small networks, allowing a theoretical characterization, which becomes very heavy and unfeasible for large networks.

This suggests the pertinence of combining the two formalisms in the study of the same system, by developing two models, one (Boolean) that comprehends all the variables known to be involved in the system and the other a reduced PWA model describing only a small set of fundamental variables. The two models carry complementary information which could not have been obtained by studying only one type of model. An example of this synergy is the NFκB model described in the Introduction. In fact, this 3-dimensional (reduced) model was obtained by application of the method in [C13] to the Boolean network described in Section 3.2. The low dimensional model was then studied in Section 3.4 as a PWA model, which enabled the inference of further details such as the existence of qualitatively distinct periodic orbits that depend on the set of parameters.

The two formalisms can complement each other in several other ways, for instance, by defining the activity functions (or production terms) of the PWA system through the Boolean logical regulatory functions at each node (as in Sections 3.1, 3.2). At the level of the state transition graphs, the PWA parameters can be used to assign probabilities of transition to each edge of the Boolean model (as in Sections 3.3, 3.4).

6.3 Predictive analysis

The mathematical methods presented here have evolved to address specific problems in biological regulatory networks, they have become more informative and predictive, to include both quantitative and qualitative aspects with a reduced amount of mathematical machinery.

One of the highlights is the characterization of the asymptotic behavior of a large (Boolean) network as a composition of two modules. It is a quite strong result, since it starts with information on the two modules only, together with the interconnecting function, but is able to predict the full asymptotic behavior of the large network. There are, of course, some limitations, such as the complexity of the algorithm as the number of outputs/inputs increases, or the decomposition of the large network into several modules, but it remains a powerful concept.

Another highlight is the idea of assigning probabilities of transition to the asynchronous transition graph based on the parameters of the PWA system. This leads to the prediction of the dynamical behavior of the system in terms of the parameters. In an example, observation of the asynchronous transition graph shows that qualitatively distinct periodic orbits may arise in the continuous system, depending on the set of parameters. It is an interesting point that this orbit can be predicted in a probabilistic way, and remarkable in the sense that not many methods exist for studying the existence of periodic orbits in a general nonlinear system. This also illustrates the strength of combining different frameworks and methodologies to complement each other, thus deriving the most knowledge from the available data and predicting new properties.

6.4 Perspectives

The methods developed throughout this *memoir* suggest a combination of modeling approaches to gain otherwise possibly inaccessible knowledge on a physical system from several sources. Some prospective applications as well as research directions are discussed below, covering different topics.

As it becomes clear from the papers assembled here, different mathematical models of a system can be useful to characterize different properties of the system. One of the questions that can then be raised is how to

construct models of the system in each of the formalisms and how to conjugate them to obtain new information.

New applications of these formalisms include techniques for constructing large networks by assembling several smaller modules (see, for instance, [C1]) or, conversely, suggest model reduction techniques. The latter concern a wide topic related to graph analysis and partition.

Lastly, these formalisms are advantageous to deal with problems concerning the regulation and control of genetic networks [C6], by suggesting and testing different control strategies.

A methodology for analysis of biological networks A general methodology would be composed of three to four steps to establish a 'hierarchy' of models that comprises several levels of abstraction (see also [C2] for a recent example).

The first step would be to start at the more abstract level, using experimental data and observations to establish a network of interactions and write a set of logical rules, which can include a finite number of discrete states for each variable. The discrete model can then be transformed into a Boolean model, by appropriately extending the state space. The Boolean model is characterized by an asynchronous transition matrix, which can itself be analyzed using tools from graph theory (such as algorithms for decomposition into strongly connected components, hierarchical organization, and attractor identification). This yields qualitative information on the dynamical behavior of the system induced by the network topology.

The second step would be to generate a piecewise affine model from the discrete model, by assigning thresholds such that each discrete state corresponds to a regular domain; the discrete rules could be written in terms of sums of products of step functions, by assigning appropriate parameters. The PWA model can be studied using analytical tools (for example, in each domain the solutions can be explicitly computed), and refines the information given by the discrete model. In addition, PWA models have a continuous state space and a set of fundamental parameters, which can be more easily estimated from experimental data. While the translation from PWA to discrete models can be well defined (as in [C11]), the reverse process is not so clear and needs to be further precised, and there may be several solutions.

The third step would be to generate a more detailed, "fully" continuous, ODE model. One immediate approach is to use Hill functions to replace step functions, to avoid the solution problems generated at the regions of vector field discontinuities, but there are clearly many possibilities open here.

From continuous to Boolean models The procedure outlined in the previous paragraph can also be used in the inverse order, that is, starting from an ODE model, successively consider more abstract modeling frameworks. Indeed, a sketch of this can be seen in the Introduction for the schematic NF-κB example. For some systems it may be easier or more appropriate to construct a first model not only from the structure of interactions, but by taking into account all details, such as the binding of two molecules, complex formation, and conservations of mass.

It is clear that a continuous model would also benefit from an analysis with rigorous theoretical tools such as those developed for discrete systems, if only there is a way to "translate" the continuous ODE into a Boolean model. The application of these qualitative methods to continuous systems, if an appropriate correspondence has been established, will be useful for finding regions where a given qualitative behavior is expected and can help with parameter estimation. An obvious question is thus how to obtain a set of logical rules that reasonably describe the ODE model and its dynamical behavior. A straightforward idea is to partition the state space into a grid and use the continuous model to assign a value to each domain or point of the grid. But there are, of course, a wide range of questions related to such procedure: how to choose the grid, given the parameters and scales of the system; how to discretize the continuous system to obtain a discrete set of values for each variables; how to assign these values at each point of the grid; how to write a set of logical rules; etc. Coupling Boolean frameworks with other techniques inspired by continuous dynamical systems may lead to improved and more intuitive methodologies.

Model reduction, composition/decomposition of networks The Boolean interconnection of two or more models is a valuable and potentially cost-effective way to construct a model of a large network and analyze its asymptotic behavior. One of the remaining questions is, however, related to the possible occurrence of spurious attractors. One first result to help decide whether an attractor of the asymptotic graph is spurious has been provided in [C3], and subsequently refined [C1]. It would be quite useful to improve or develop new results to decrease the gap between the outcome of the asymptotic graph (which gives all the attractors of the large model, but possibly also some spurious) and the real system.

On the other hand, given a large network, and knowing that the size of a transition graph grows exponentially with the network's dimension, it becomes a pertinent problem to find appropriate techniques to decompose a network into two or more modules. This is known to be a NP-hard problem in graph-partitioning, but some classical algorithms that minimize the edges between modules exist, namely spectral techniques [68], hierarchical clustering [39] and Markov cluster algorithm [13]. These have been used in [C3], to illustrate the model reduction idea. However, a critical point will be to balance the number and size of the modules with the number of inputs (or edges) connecting each of the modules, in order to increase the effectiveness of such a model reduction technique.

Control design The study of control problems generally requires three points to be known: a model of the system, the outputs or measured variables, which give information on the state of the system over time, and the inputs or control parameters, which allow a user to manipulate the system. The classical procedures for control design are not well adapted for application to biological molecular systems, where the measured variables are typically the concentration or expression levels of a group of proteins and/or mRNAs. A major difficulty is the time frequency of data sampling: instead of having access to continuous, online, measurements, it is more likely that we have to deal with qualitative indications of protein expression levels at spaced intervals. Another difficulty is the range of values and functional forms allowed for the inputs. Some of the possible inputs include rates of transcription of certain genes (for instance, through the construction of plasmids containing inducers to those genes [27], so that the transcription rate of a gene can be increased by a certain factor simply by adding a suitable amount of the corresponding inducer molecules to the system). With the plasmid experimental techniques, it may be difficult to tune the amount of inducer to obtain a very precise transcription rate, so the control range is again limited to qualitative indications.

Therefore, by their qualitative nature, the methods described throughout this *memoir* seem appropriate to adapt and develop control strategies that can deal with the various constraints imposed by the biological system. Even though details may be scarce and information only partially available, discrete or PWA frameworks can still suggest qualitative strategies for control to a desired state ([20],[C9]). Most importantly, the design of feedback control laws must take into account the experimental setup and an implementation [60, 67] using biological components or by re-wiring the network through feasible interconnections. This is the goal, for instance, in the projects GeMCo (ANR) and RESET (Investissements d'Avenir) (see Section **??**), which are dedicated to several control objectives for the growth rate of bacteria *E. coli* using synthetic circuit engineering.

Bibliography

[1] R. Albert and H. G. Othmer. The topology of the regulatory interactions predicts the expression pattern of the *Drosophila* segment polarity genes. *J. Theor. Biol.*, 223:1–18, 2003.

[2] U. Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall/CRC, Boca Raton, 2006.

[3] E. Andrianantoandro, S. Basu, D. Karig, and R. Weiss. Synthetic biology: new engineering rules for an emerging discipline. *Mol. Syst. Biol.*, 2:0–0, 2006.

[4] D. Angeli, J.E. Ferrell, and E.D. Sontag. Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *PNAS*, 101(7):1822–1827, 2004.

[5] D. Angeli and E.D. Sontag. Monotone control systems. *IEEE Trans. Automat. Control*, 48(10):1684 – 1698, 2003.

[6] S. Azuma, E. Yanagisawa, and J. Imura. Controllability analysis of biosystems based on piecewise affine systems approach. *IEEE Trans. Automat. Control*, 53:139–152, 2008.

[7] A. Becskei and L. Serrano. Engineering stability in gene networks by autoregulation. *Nature*, 405:590–593, 2000.

[8] G. Bernot, J.-P. Comet, A. Richard, and J. Guespin. Application of formal methods to biological regulatory networks: Extending Thomas' asynchronous logical approach with temporal logic. *Journal of Theoretical Biology*, 229(3):339–347, 2004.

[9] L. Calzone, L. Tournier, S. Fourquet, D. Thieffry, B. Zhivotovsky, E. Barillot, and A. Zinovyev. Mathematical modelling of cell-fate decision in response to death receptor engagement. *PLoS Comput. Biol.*, 6(3):e1000702, 2010.

[10] R. Casey, H. de Jong, and J.L. Gouzé. Piecewise-linear models of genetic regulatory networks: equilibria and their stability. *J. Math. Biol.*, 52:27–56, 2006.

[11] M I Davidich and S Bornholdt. Boolean network model predicts cell cycle sequence of fission yeast. *PLoS ONE*, 3(2):e1672, 02 2008.

[12] H. de Jong, J. Geiselmann, C. Hernandez, and M. Page. Genetic network analyzer: qualitative simulation of genetic regulatory networks. *Bioinformatics*, 19:336–344, 2003.

[13] S. Van Dongen. Graph clustering via a discrete uncoupling process. *SIAM J. Matrix Analysis and Applications*, 30:121141, 2008.

[14] L. Edelstein-Keshet. *Mathematical models in Biology*. SIAM classics in applied mathematics, Philadelphia, 2005.

[15] R. Edwards. Analysis of continuous-time switching networks. *Physica D*, 146:165–199, 2000.

[16] R. Edwards and L. Glass. Combinatorial explosion in model gene networks. *Chaos*, 10:691–704, 2000.

[17] T. Eissing, H. Conzelmann, E.D. Gilles, F. Allgöwer, E. Bullinger, and P. Scheurich. Bistability analysis of a caspase activation model for receptor-induced apoptosis. *J. Biol. Chem.*, 279:36892–36897, 2004.

[18] M.B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403:335–338, 2000.

[19] F. Fages, S. Soliman, and N. Chabrier-Rivier. Modelling and querying interaction networks in the biochemical abstract machine BIOCHAM. *J. Biological Physics and Chemistry*, 4(2):64–73, 2005.

[20] E. Farcot and J.L. Gouzé. A mathematical framework for the control of piecewise-affine models of gene networks. *Automatica*, 44(9):2326–2332, 2008.

[21] E. Farcot and J.L. Gouzé. Periodic solutions of piecewise affine gene network models with non uniform decay rates: The case of a negative feedback loop. *Acta Biotheoretica*, 57(4):429–455, 2009.

[22] A. Fauré, A. Naldi, C. Chaouiya, and D. Thieffry. Dynamical analysis of a generic boolean model for the control of the mammalian cell cycle. *Bioinformatics*, 22(14):e124–e131, 2006.

[23] M. Feinberg. Mathematical aspects of mass action kinetics. In L. Lapidus and N. Amundson, editors, *Chemical Reactor Theory: A Review*, pages 1–78. Prentice-Hall, Englewood Cliffs, 1977.

[24] M. Feinberg. Chemical reaction network structure and the stabiliy of complex isothermal reactors - i. the deficiency zero and deficiency one theorems. *Chemical Engineering Science*, 42:2229–2268, 1987.

[25] W. Feller. *An introduction to probability theory and its applications*. Singapore: Wiley, 1970.

[26] A.F. Filippov. *Differential equations with discontinuous righthand-sides*. Kluwer Academics Publishers, 1988.

[27] T.S. Gardner, C.R. Cantor, and J.J. Collins. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 403:339–342, 2000.

[28] F. Giannakopoulos and K. Pliete. Planar systems of piecewise linear differential equations with a line of discontinuity. *Nonlinearity*, 14(6):16111632, 2001.

[29] L. Glass. Combinatorial and topological methods in nonlinear chemical kinetics. *J. Chem. Phys.*, 63:1325–1335, 1975.

[30] L. Glass and S.A. Kauffman. The logical analysis of continuous, nonlinear biochemical control networks. *J. Theor. Biol.*, 39:103–129, 1973.

[31] L. Glass and J.S. Pasternak. Stable oscillations in mathematical models of biological control systems. *J. Math. Biol.*, 6:207–223, 1978.

[32] A.G. Gonzalez, A. Naldi, L. Sànchez, D.Thieffry, and C. Chaouiya. GINsim: a software suite for the qualitative modelling, simulation and analysis of regulatory networks. *BioSystems*, 84(2):91–100, 2006.

[33] J.L. Gouzé and T. Sari. A class of piecewise linear differential equations arising in biological models. *Dyn. Syst.*, 17(4):299–316, 2002.

[34] L.C.G.J.M. Habets and J. van Schuppen. A control problem for affine dynamical systems on a full-dimensional polytope. *Automatica*, 40:21–35, 2004.

[35] R. Heinrich and S. Schuster. *The regulation of cellular systems*. Springer, Berlin, 1996.

[36] A. Hoffmann, A. Levchenko, M.L. Scott, and D. Baltimore. The IkB-NFkB signaling module: temporal control and selective gene activation. *Science*, 298:1241–1245, 2002.

[37] F. Horn and R. Jackson. General mass action kinetics. *Archive for Rational Mechanics and Analysis*, 47:81–116, 1972.

[38] S. Jamshidi, H. Siebert, and A. Bockmayr. Comparing discrete and piecewise affine differential equation models of gene regulatory networks. In M. Lones, S. Smith, S. Teichmann, F. Naef, J. Walker, and M. Trefzer, editors, *Information Processign in Cells and Tissues*, volume 7223 of *Lecture Notes in Computer Science*, pages 17–24. Springer Berlin / Heidelberg, 2012.

[39] S.C. Johnson. Hierarchical clustering schemes. *Psychometrika*, 32:241254, 1967.

[40] H. De Jong. Modeling and simulation of genetic regulatory systems: a literature review. *Journal of Computational Biology*, 9(1):67–103, 2002.

[41] J. Keener and J. Sneyd. *Mathematical Physiology*. Springer-Verlag, New York, 1998.

[42] H.K. Khalil. *Nonlinear systems*. Prentice Hall, New Jersey, 2002.

[43] H. Kitano. Biological robustness. *Nature Reviews Genetics*, 5:826, 2004.

[44] E. Klipp, R. Herwig, A. Howald, C. Wierling, and H. Lehrach. *Systems Biology in Practice*. Wiley-VCH, Weinheim, 2005.

[45] F. Li, T. Long, Y. Lu, Q. Ouyang, and C. Tang. The yeast cell-cycle network is robustly designed. *PNAS*, 101:47814786, 2004.

[46] S. Li, S.M. Assmann, and R. Albert. Predicting essential components of signal transduction networks: A dynamic model of guard cell abscisic acid signaling. *PLoS Biol*, 4(10):e312, 09 2006.

[47] T. Lipniacki, P. Paszek, A.R. Brasier, B. Luxon, and Marek Kimmel. Mathematical model of NFκB regulatory module. *J. Theor. Biol.*, 228:195–215, 2004.

[48] J.A. Papin, T. Hunter, B.O. Palsson, and S. Subramaniam. Reconstruction of cellular signalling networks and analysis of their properties. *Nature Rev. Mol. Cell. Biol.*, 6(2):99–111, 2005.

[49] E. Remy, P. Ruet, and D. Thieffry. Graphic requirements for multistability and attractive cycles in a Boolean dynamical framework. *Advances in Applied Mathematics*, 41(3), 2008.

[50] A. Richard. Positive circuits and maximal number of fixed points in discrete dynamical systems. *Discrete Applied Mathematics*, 157(15):3281–3288, 2009.

[51] A. Richard. Negative circuits and sustained oscillations in asynchronous automata networks. *Advances in Applied Mathematics*, 44(4):378–392, 2010.

[52] D. Ropers, H. de Jong, M. Page, D. Schneider, and J. Geiselmann. Qualitative simulation of the carbon starvation response in *Escherichia coli*. *Biosystems*, 84(2):124–152, 2006.

[53] J Saez-Rodriguez, L Simeoni, J A Lindquist, R Hemenway, U Bommhardt, B Arndt, U-U Haus, R Weismantel, E D Gilles, S Klamt, and B Schraven. A logical model provides insights into t cell receptor signaling. *PLoS Comput Biol*, 3(8):e163, Aug 2007.

[54] L. Sánchez and D. Thieffry. A logical analysis of the *drosophila* gap-gene system. *J. Theor. Biol.*, 211:115–141, 2001.

[55] R. Schlatter, K. Schmich, I. Avalos Vizcarra, P. Scheurich, T. Sauter, et al. ON/OFF and beyond - a Boolean model of apoptosis. *PLoS Comput. Biol.*, 5(12):e1000595, 2009.

[56] V. Sevim, X. Gong, and J.E. Socolar. Reliability of transcriptional cycles and the yeast cell-cycle oscillator. *PLoS Comput. Biol.*, 6:e1000842, 2010.

[57] H. Siebert and A. Bockmayr. Temporal constraints in the logical analysis of regulatory networks. *Theoretical Computer Science*, 391:258–275, 2008.

[58] M. Smoot, K. Ono, J. Ruscheinski, P.-L. Wang, and T. Ideker. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, 27(3):431432, 2011.

[59] E.D. Sontag. *Mathematical Control Theory (2nd ed.)*. Springer-Verlag, New York, 1998.

[60] E.D. Sontag. Some new directions in control theory inspired by systems biology. *Syst. Bio.*, 1(1):1–18, 2004.

[61] R. Thomas. Boolean formalization of genetic control circuits. *J. Theor. Biol.*, 42:563–585, 1973.

[62] R. Thomas and R. D'Ari. *Biological feedback*. CRC Press, 1990.

[63] R. Thomas and M. Kaufman. Multistationarity, the basis of cell differentiation and memory. I. Structural conditions of multistationarity and other nontrivial behavior. *Chaos*, 11(1):170–179, 2001.

[64] M. Tigges, T.T. Marquez-Lago, J. Stelling, and M. Fussenegger. A tunable synthetic mammalian oscillator. *Nature*, 457:309–312, 2009.

[65] P. van Ham. How to deal with more than two levels. In R. Thomas, editor, *Kinetic Logic: A Boolean Approach to the Analysis of Complex Regulatory Systems*, volume 29 of *Lecture Notes in Biomathematics*, pages 326–343. Springer, 1979.

[66] D. Del Vecchio, A.J. Ninfa, and E.D. Sontag. Modular cell biology: Retroactivity and insulation. *Molecular Systems Biology*, 4:161, 2008.

[67] D. Del Vecchio and E.D. Sontag. Synthetic biology: A systems engineering perspective. In P.A. Iglesias and B.P. Ingalls, editors, *Control Theory and Systems Biology*, pages 101–124. MIT Press, 2009.

[68] U. von Luxburg. A tutorial on spectral clustering. *Stat Comput*, 17:395–416, 2007.

[69] R. S. Wang, A. Saadatpour, and R. Albert. Boolean modeling in systems biology: an overview of methodology and applications. *Physical Biology*, 9:055001, 2012.

[70] G. Yagil and E. Yagil. On the relation between effector concentration and the rate of induced enzyme synthesis. *Biophys. J.*, 11:11–27, 1971.

[71] R. Zhang, M.V. Shah, J. Yang, S.B. Nyland, X. Liu, J.K. Yun, R. Albert, and T.P. Loughran Jr. Network model of survival signaling in LGL leukemia. *PNAS*, 105:16308–16313, 2008.

Appendix A

List of publications

A.1 Journal articles, book chapters and thesis

1. M. Chaves and A. Carta. Attractor computation using interconnected Boolean networks: testing growth rate models in *E. Coli*. *Theoretical Computer Science*, submitted.

2. M. Chaves and M. Preto. Hierarchy of models: from qualitative to quantitative analysis of circadian rhythms in cyanobacteria. *Chaos*, 23(2):025113, 2013.

3. L. Tournier and M. Chaves. Interconnection of asynchronous boolean networks, asymptotic and transient dynamics. *Automatica*, 49:884-893, 2013.

4. M. Chaves, E. Farcot, and J.-L. Gouzé. Probabilistic approach for predicting periodic orbits in piecewise affine differential models. *Bull. Math. Biol.*, 75(6):967–987, 2013.

5. G. Bernot, J.-P. Comet, A. Richard, M. Chaves, J.-L. Gouzé, and F. Dayan. Modeling and analysis of gene regulatory networks. In F. Cazals and P. Kornprobst, editors, *Modeling in Computational Biology and Biomedicine*, pages 47–80. Springer-Verlag Heidelberg, 2013.

6. A. Carta, M. Chaves, and J.-L. Gouzé. A simple model to control growth rate of synthetic E. coli during the exponential phase: model analysis and parameter estimation. In D. Gilbert and M. Heiner, editors, *CMSB 2012*, Lecture Notes in Computer Science 7605, pages 107-126. Springer, 2012. [1]

7. D.A. Oyarzún, M. Chaves, and M. Hoffmeyer-Zlotnik. Multistability and oscillations in genetic control of metabolism. *J. Theor. Biol.*, 295:139–153, 2012.

8. W. Abou-Jaoudé, M. Chaves, and J. L. Gouzé. A theoretical exploration of birhythmicity in the p53-mdm2 network. *PLoS ONE*, 6(2):e17075, 2011.

9. M. Chaves and J.L. Gouzé. Exact control of genetic networks in a qualitative framework: the bistable switch example. *Automatica*, 47:1105–1112, 2011.

10. I. Ndiaye, M. Chaves, and J. L. Gouzé. Oscillations induced by different timescales in signal modules regulated by slowly evolving protein-protein interactions. *IET Systems Biology*, 4(4):263–276, 2010.

11. M. Chaves, L. Tournier, and J. L. Gouzé. Comparing Boolean and piecewise affine differential models for genetic networks. *Acta Biotheoretica*, 58(2):217–232, 2010.

[1] This article won the *Best Student Paper Award* (Alfonso Carta) at the $10th$ Conference in Computational Methods in Systems Biology (2012).

12. J. L. Gouzé and M. Chaves. Piecewise affine models of regulatory genetic networks: review and probabilistic interpretation. In J. Lévine and P. Müllhaupt, editors, *Advances in the Theory of Control, Signals and Systems, with Physical Modelling*, volume 470 of *Lecture Notes in Control and Information Sciences*, pages 241–253. Springer, 2010.

13. L. Tournier and M. Chaves. Uncovering operational interactions in genetic networks using asynchronous boolean dynamics. *J. Theor. Biol.*, 260(2):196–209, 2009.

14. T. Eissing, M. Chaves, and F. Allgöwer. Live and let die - a systems biology view on cell death. *Computers & Chemical Engineering*, 33(3):583–589, 2009.

15. M. Chaves, T. Eißing, and F. Allgöwer. Regulation of apoptosis via the NFκB pathway: modeling and analysis. In A. Deutsch N. Ganguly and A. Mukherjee, editors, *Dynamics on and of complex networks: applications to biology, computer science and the social sciences*, Modeling and Simulation in Science, Engineering and Technology, pages 19–34. Birkhauser, Boston, 2009.

16. M. Chaves, A. Sengupta, and E.D. Sontag. Geometry and topology of parameter space: investigating measures of robustness in regulatory networks. *J. Math. Biol.*, 59(3):315–358, 2009.

17. A. Dayarian, M. Chaves, A. Sengupta, and E.D. Sontag. Shape, size and robustness: feasible regions in the parameter space of biochemical networks. *PLoS Comp. Biol.*, 5(1):e1000256, 2009.

18. M. Chaves and R. Albert. Studying the effect of cell division on expression patterns of the segment polarity genes. *J. Royal Society Interface*, 5(S1):S71–S84, 2008.

19. M. Chaves, T. Eissing, and F. Allgöwer. Bistable biological systems: a characterization through local compact input-to-state stability. *IEEE Trans. Automatic Control*, 53(1):87–100, 2008.

20. E.D. Sontag and M. Chaves. Exact computation of amplification for a class of nonlinear systems arising from cellular signaling pathways. *Automatica*, 42(11):1987–1992, 2006.

21. M. Chaves, E.D. Sontag, and R. Albert. Methods of robustness analysis for boolean models of gene control networks. *IEE Proc. Syst. Biol.*, 153:154–167, 2006.

22. M. Chaves. Input-to-state stability of rate-controlled biochemical networks. *SIAM J. Control and Optimization*, 44(2):704–727, 2005.

23. M. Chaves, R. Albert, and E.D. Sontag. Robustness and fragility of boolean models for genetic regulatory networks. *J. Theor. Biol.*, 235:431–449, 2005. [2]

24. M. Chaves, E.D. Sontag, and R.J. Dinerstein. Steady-states of receptor–ligand dynamics: a theoretical framework. *J. Theor. Biol.*, 227(3):413–428, 2004.

25. M. Chaves, E.D. Sontag, and R.J. Dinerstein. Optimal length and signal amplification in weakly activated signal transduction cascades. *J. Physical Chemistry B*, 108(39):15311–15320, 2004.

26. M. Chaves. *Observer design for a class of nonlinear systems, with applications to biochemical networks.* PhD thesis, Rutgers University, New Jersey, USA, 2003.

27. M. Chaves and E.D. Sontag. State-estimators for chemical reaction networks of Feinberg-Horn-Jackson zero deficiency type. *Eur. J. Control*, 8(4):343–359, 2002.

[2] This article was for several months in the "Top 10 most cited papers of the last 5 years" of the Journal of Theoretical Biology

A.2 Peer-reviewed conference proceedings

28. C. Breindl, M. Chaves, and Frank Allgöwer. . A linear reformulation of Boolean optimization problems and its application to the problem of estimating the structure of gene regulation networks. *Proc.* 52^{th} *Conf. Decision and Control*, Florence, Italy, December 2013, to appear.

29. A. Carta, M. Chaves, and J.-L. Gouzé. A class of switched piecewise quadratic systems for coupling gene expression with growth in bacteria. *Proc.* 9^{th} *IFAC Symp. Nonlinear Control Systems (NOLCOS'13)*, Toulouse, France, September 2013, to appear.

30. C. Breindl, M. Chaves, J.-L. Gouzé, and Frank Allgöwer. Structure estimation for unate Boolean models of gene regulation networks. In *Proc.* 16^{th} *IFAC Symposium on System Identification*, Brussels, Belgium, July 2012.

31. X.-D. Li, M. Chaves, and J.-L. Gouzé. Robust estimation for hybrid models of genetic networks. In *Proc.* 20^{th} *Mediterranean Conf. on Control and Automation*, Barcelona, Spain, July 2012.

32. M. Chaves and L. Tournier. Predicting the asymptotic dynamics of large biological networks by interconnections of boolean modules. In *Proc.* 50^{th} *Conf. Decision and Control and European Control Conf.*, Orlando, Florida, USA, December 2011.

33. X.-D. Li, J.L. Gouzé, and M. Chaves. An observer for a genetic network model with boolean observations. In *Proc.* 50^{th} *Conf. Decision and Control and European Control Conf.*, Orlando, Florida, USA, December 2011.

34. D. A. Oyarzún and M. Chaves. Global gene regulation in metabolic networks. In *Proc.* 18^{th} *IFAC World Congress*, Milan, Italy, August 2011.

35. W. Abou-Jaoudé, M. Chaves, and J. L. Gouzé. Mechanisms for coexistence of two limit cycles in a biochemical model. In *Proc.* 18^{th} *IFAC World Congress*, Milan, Italy, August 2011.

36. M. Chaves and J. L. Gouzé. Qualitative control of genetic networks: the bistable switch example. In *Proc.* 8^{th} *IFAC Symp. Nonlinear Control Systems (NOLCOS'10)*, Bologna, Italy, September 2010.

37. M. Chaves, E. Farcot, and J. L. Gouzé. Transition probabilities for piecewise affine models of genetic networks. In *Proc.* 19^{th} *Int. Symp. Mathematical Theory of Networks and Systems (MTNS'10)*, Budapest, Hungary, July 2010.

38. M. Chaves. Methods for qualitative analysis of genetic networks. In *Proc.* 10^{th} *European Control Conf. (ECC'09)*, pages 671–676. Budapest, Hungary, August 2009.

39. L. Tournier and M. Chaves. Operational interactions in genetic networks: application to an apoptosis signalling pathway. In *Proc.* 10^{th} *European Control Conf. (ECC'09)*, pages 1889–1894. Budapest, Hungary, August 2009.

40. I. Ndiaye, M. Chaves, and J.L. Gouzé. Study and parameter identification of a model coupling cell signaling and gene expression. In *Proc.* 16^{th} *Mediterranean Conf. Control and Automation (MED'08)*. Ajaccio, France, June 2008.

41. I. Ndiaye, M. Chaves, and J.L. Gouzé. Un petit modèle d'interaction entre expression genétique et signalisation. In *Réseaux d'interactions : analyse, modélisation et simulation*, Integrative Post-Genomics, Lyon, France, November 2007.

42. S. Waldherr, T. Eissing, M. Chaves, and F. Allgöwer. Bistability preserving model reduction in apoptosis. In *Proc.* 10^{th} *IFAC Symp. on Computer Applications in Biotechnology (CAB'07)*, Cancun, Mexico, pages 327–332, June 2007.

43. M. Chaves, E.D. Sontag, and R. Albert. Structure and timescale analysis in genetic regulatory networks. In *Proc.* 45^{th} *Conf. Decision and Control (CDC'06)*, San Diego, CA, USA, pages 2358–2363, December 2006.

44. M. Chaves. Stability of rate-controlled zero-deficiency networks. In *Proc.* 45^{th} *Conf. Decision and Control (CDC'06)*, San Diego, CA, USA, pages 5766–5771, December 2006.

45. M. Chaves, T. Eißing, and F. Allgöwer. Identifying mechanisms for bistability in an apoptosis network. In *Réseaux d'interactions : analyse, modélisation et simulation*, Integrative Post-Genomics, Lyon, France, November 2006.

46. E.D. Sontag and M. Chaves. Computation of amplification for systems arising from cellular signaling pathways. In *Proc.* 16^{th} *IFAC World Congress*, Prague, Czech Republic, July 2005.

47. M. Chaves, E.D. Sontag, and R.J. Dinerstein. Gains and optimal design in signaling pathways. In *Proc.* 43^{th} *Conf. Decision and Control (CDC'04)*, Paradise Island, The Bahamas, December 2004.

48. S. Glavaski, M. Chaves, R. Day, P. Nag, A. Wiliams, and W. Zhang. Vehicle networks: achieving regular formation. In *Proc. American Control Conf. (ACC'03)*, Denver, Colorado, June 2003.

49. M. Chaves. A parameter-robust observer as an application of iss techniques. In *Proc.* 15^{th} *Int. Symp. Mathematical Theory of Networks and Systems (MTNS'02)*, South Bend, Indiana, USA, August 2002.

50. M. Chaves and E.D. Sontag. Observers for chemical reaction networks. In *Proc.* 6^{th} *European Control Conf. (ECC'01)*, Porto, Portugal, September 2001.

51. M. Chaves and E.D. Sontag. An alternative observer for zero deficiency chemical networks. In *Proc.* 5^{th} *IFAC Symp. Nonlinear Control Systems (NOLCOS'01)*, St. Petersburg, Russia, July 2001.

A.3 Technical reports

52. M. Chaves, E.D. Sontag, and A. Sengupta. Shape, size and robustness: feasible regions in the parameter space of biochemical networks. Technical Report arXiv:0710.4269v1, q-bio.MN, http://arXiv.org, 2007.

53. M. Chaves, R. Day, L. Gomez-Ramos, P. Nag, A. Wiliams, W. Zhang, and S. Glavaski. Vehicle networks: achieving regular formation. Technical report. Mathematical modeling in industry – IMA Summer program for graduate students, May 26-June 3 2002 (R. Kuske, R. Reitich and F. Santosa, organizers). Institute for Mathematics and its Applications, Preprint 1866, paper 1866-2.

Appendix B

Collected articles (original publications)

NOTE: The original articles have been removed from this version of the manuscript. Some of them are freely accessible. Preliminary versions can also be found in my webpage, `http://www-sop.inria.fr/members/Madalena.Chaves`.

The book chapter:

"Modeling and analysis of gene regulatory networks" by G. Bernot, J.-P. Comet, A. Richard, M. Chaves, J.-L. Gouzé, and F. Dayan. In F. Cazals and P. Kornprobst, editors, *Modeling in Computational Biology and Biomedicine*, pages 47–80. Springer-Verlag Heidelberg, 2013.

is available for download, as sample pages, from the Springer's website `http://www.springer.com/new+%26+forthcoming+titles+28default%29/book/978-3-642-31207-6`.